Mechanisms of multipass membrane protein biogenesis

Patrick J. Chitwood

MRC Laboratory of Molecular Biology
and
St John's College

This dissertation is submitted for the degree of

Doctor of Philosophy

University of Cambridge
March 2020

I hope reading this thesis engenders the same exhilaration and curiosity I experienced when first seeing the results.

Declaration

This dissertation is the result of my own work and includes nothing which is the work done in collaboration except where specifically indicated in the figure legends. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Biology Degree Committee.

Patrick J. Chitwood,
March 2020

Summary

Mechanisms of multipass membrane protein biogenesis

Patrick J. Chitwood

A crucial aspect of cellular physiology is the ability of a single cell to remain autonomous and concentrate reagents to improve efficiency. Semi-permeable membranes facilitate autonomy via an outer barrier (plasma membrane) and enclose functional hubs (organelles) to efficiently carry out biological processes. However, cells do not live in isolation and must communicate with neighboring cells, uptake and traffic nutrients, and react to a dynamic extracellular environment. All these processes require integral membrane proteins (IMPs), which are embedded within all cellular membranes. Highlighting their importance is a myriad of human diseases observed upon disruption of their biogenesis. This thesis aims to describe our recent contributions to the understanding of how membrane proteins are made at the endoplasmic reticulum (ER), the primary site of IMP biogenesis in the cell.

Although IMPs are defined by a single feature, a transmembrane domain (TMD), the ~5000 encoded in the mammalian genome are diverse. While some contain a single TMD, most feature many biophysically unique TMDs. Productive biogenesis of most "multipass" membrane proteins requires insertion in a defined topology as well as packing of their TMDs into a helical bundle, often through interactions between polar residues unstably located in the hydrophobic lipid bilayer. Neither the mechanisms that facilitate accurate topological insertion or the subsequent stabilization of polar TMDs during IMP biogenesis are completely understood.

First, we demonstrate the efficient topogenesis of many GPCRs requires the conserved ER membrane protein complex (EMC). This is supported by biochemical reconstitution of β1-adrenergic receptor (β_1AR) insertion in-vitro, which placed EMC at an early step during co-translational insertion of the first TMD (TMD1). In the absence of EMC, TMD1 was topologically inverted or failed to insert altogether. EMC and SRP receptor were sufficient for the correct insertion of TMD1, while

insertion of the next TMD required Sec61. Finally, EMC necessity could be bypassed by enforcement of TMD1 topology via an N-terminal signal peptide. Following accurate insertion of TMD1, we define the engagement of a newly identified intramembrane chaperone protein complex that we term the PAT complex. The PAT complex is an obligate heterodimer consisting of the highly conserved proteins CCDC47 and Asterix. A diverse set of multipass membrane proteins show impaired biogenesis upon PAT complex depletion, despite correct topological insertion. Biochemical analyses demonstrate PAT complex engages nascent TMDs that contain unshielded polar amino acids but disengages upon substrate folding. Thus, EMC cooperates with Sec61 to co-translationally insert TMDs, ensuring accurate membrane protein topogenesis, while the PAT complex acts after insertion to protect transmembrane domains during their assembly.

Publications

Chitwood, P.J. and Hegde, R.S. (2019). The role of EMC during Membrane Protein Biogenesis. Trends Cell Biol, 29 (5), 371-384.

Chitwood, P.J., Juszkiewicz, S**.**, Guna, A., Shao, S., and Hegde, R.S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. Cell 175, 1507–1519.e16.

Acknowledgements

The journey through the PhD is an incredible one, filled with excitement and fulfilment upon making an initial discovery and finally understanding how a biological process works. Experiencing this was only possible with great support from family, friends, colleagues, and a great mentor.

The best supporters of this process have been my mother Chris and father Dan. Though not familiar with science or academia, they always encouraged pursuing a career that I am passionate about and fully believed in the importance of this endeavor. I would also like to thank my brother Drew and sister Kate, who are not only my siblings but also my best friends and a constant source of support. My family was the primary driving force that persuaded me to move half-way across the world and constantly remind me that the difficult things are the most worthwhile. This sentiment is best embodied by the great life and friendships I was able to build during my time at Cambridge. I would like to especially thank Phillip Lindstedt, Tom Mosley, Fynn McKay, Gaelen Stanford-Moore, and Lisa Hawkins. They have become lifelong friends and provided me with some of the best memories outside the laboratory.

Early in my scientific career I was advised on the importance of good colleagues and a good mentor. This is now more evident than ever as my time at the LMB comes to an end. I don't believe I will ever find another place with such a collective passion for science. This was especially true in the Hegde lab. My favorite memories will always be the long scientific discussions with Szymon Juszkiewicz, Alina Guna, Eszter Zavodsky, Zhewang Lin, and John O'Donnel. Finally, I would like to thank my wonderful mentor Manu. I came to his lab with the narrow goal of learning mechanistic biochemistry, but I am leaving with a broader outlook and better understanding of how to pursue discovery-based science. His training has engendered a mindset towards mechanistic understanding and rigorously pursuing a scientific question until one arrives at the correct answer. For this, I will be forever grateful.

Table of Contents

III An intramembrane chaperone facilitates multipass membrane protein biogenesis

Chapter I

Membrane protein insertion

The membrane proteome

Approximately 20-30% of the eukaryotic proteome is composed of proteins which are embedded within the hydrophobic environment of a lipid bilayer (Fagerberg et al., 2010; Kumar et al., 2002). Termed integral membrane proteins (IMPs), this important class is essential in many biological processes such as cell signaling, intracellular trafficking, inter- or intra-cellular communication, organelle biogenesis, and many more. The minimal requirement for integration of an IMP is a single transmembrane domain (TMD): which optimally consists of a continuous stretch of 18-28 predominantly hydrophobic residues adopting an alpha-helical secondary structure (Argos et al., 1982; Engelman et al., 1986). All IMPs in the nuclear encoded genome must be synthesized by cytosolic ribosomes, inserted into the lipid bilayer and targeted to their appropriate membranes. The inherent necessity of cytosolic translation poses several biophysical challenges to the productive biogenesis of IMPs. First, stretches of hydrophobic polypeptide are energetically unfavorable in the aqueous cytosol and prone to off-pathway interactions or misfolding. Second, IMPs often contain functional domains localized on either side the membrane and premature folding (or misfolding) can preclude membrane insertion. Finally, IMPs are localized throughout the cell in a myriad of specialized membrane systems, making direct targeting a logistical nightmare. These challenges are overcome by coupling synthesis and insertion, producing them co-translationally on the surface of the largest endomembrane system, the endoplasmic reticulum (ER). Co-translational integration minimizes exposure of hydrophobic TMDs to the aqueous cytosol and completely bypasses premature folding. Additionally, localizing biogenesis to a single membrane simplifies the targeting problem. Consistent with these energetic and logistic advantages, nearly all IMPs are produced and assembled at the ER before being trafficked to their final intracellular destinations (Figure 1). Similarly, all secretory proteins are translated and translocated into the ER lumen prior to trafficking and secretion through the exocytic system.

Few exceptions exist and are incompatible with co-translational integration. These include a small subset of proteins that elude recognition by the co-translational targeting machinery, and mitochondrial membrane proteins. Failure of co-translational TMD recognition occurs as a result of terminating translation before a

TMD has been exposed to the cytosol (Kutay et al., 1993). This phenomena is explained by the shielding of ~35-40 residues in the ribosomal exit tunnel during translation (Malkin and Rich, 1967; Sabatini and Blobel, 1970). As a result, a minimum nascent chain length of ~50-60 amino acids is required for co-translational recognition and targeting (Walter and Blobel, 1981). Thus, extremely small IMPs or those with their TMD located <60 residues from the C-terminus must be targeted and inserted into the membrane "post-translationally" (Kutay et al., 1993). While few examples of small (<70 residues) functional membrane proteins exist (Chi et al., 1996; Navarre et al., 1994; Wawrzynow et al., 1992), the latter are termed tail-anchored (TA) proteins and constitute a small but essential part of the mammalian membrane proteome (Beilharz et al., 2003). Reflecting their importance, a widely conserved set of machineries have evolved to facilitate post-translational integration into the membrane. Termed the TRC40 pathway in mammals or the GET pathway in yeast, this is the primary route for post-translational integration into the ER membrane. As such, it is equipped with dedicated targeting machineries in the cytosol and an insertase complex resident in the ER membrane (Hegde and Keenan, 2011). Similarly, mitochondrial membrane proteins (MMPs) bypass all ER dedicated targeting and insertion machinery. Not only do they require an entirely different targeting sequence, but also a set of translocases and chaperones resident in the multi-membrane system of the mitochondria (Becker et al., 2009). Thus, the key step determining both the location and machinery by which a newly synthesized IMP will be inserted is during recognition and targeting. This concept also holds true for the other two membrane insertion pathways described: peroxisomes and chloroplasts. In the case of peroxisomal membrane proteins (PMPs), a dedicated cytosolic chaperone/targeting factor recognizes PMPs and delivers them to the peroxisomal membrane via a resident targeting receptor (Hettema et al., 2014; Ma et al., 2011). Chloroplasts similarly have their own dedicated targeting and insertion machineries that are most analogous to the system observed in mitochondria (Hofmann and Theg, 2005; Kim et al., 2019). The largest mystery in regard to the coordination of membrane protein targeting in the cell is how various IMPs bypass the predominant SRP mediated targeting mechanisms, so that they can be recognized by their organelle specific targeting factors. In the case of tail-anchored proteins (present in all organelles) the

answer is rather straightforward, as the TMD is not exposed to the cytosol until translation terminates and thus never has the opportunity to be recognized by the ribosome associated SRP. However for chloroplast, mitochondria and some peroxisomal signal anchor proteins the mechanism remains unresolved. The recent discovery of a chloroplast specific targeting factor, AKR2, has shed some light on a potential mechanism. AKR2 has affinity for the ribosomal protein RPL23A, positioning it near the ribosomal exit tunnel in order to capture a chloroplast specific TMD (Kim et al., 2019). However, the exact details of how such a targeting factor would be correctly recruited to a ribosome and how it would preclude/out-compete SRP in order to achieve specificity and efficient targeting, remains entirely unclear.

Although all of these specialized membrane protein biogenesis pathways are essential for life in higher eukaryotes (Baker et al., 1990; Mukhopadhyay et al., 2006), they represent alternatives and together accommodate a relatively small percentage of IMPs: 3-5% (TA-proteins) and 6-8% (MMPs) (Kalbfleisch et al., 2007; Reinders et al., 2006 and Figure 1.1), <1% (PMPs) and chloroplasts being plant specific ((Ma et al., 2011; Schlüter et al., 2007). Thus, co-translational integration at the ER represents the primary route for membrane protein biogenesis. Its importance in human physiology is further buttressed by evidence suggesting ~50% of small molecule drugs target membrane proteins integrated via this pathway (Santos et al., 2017).

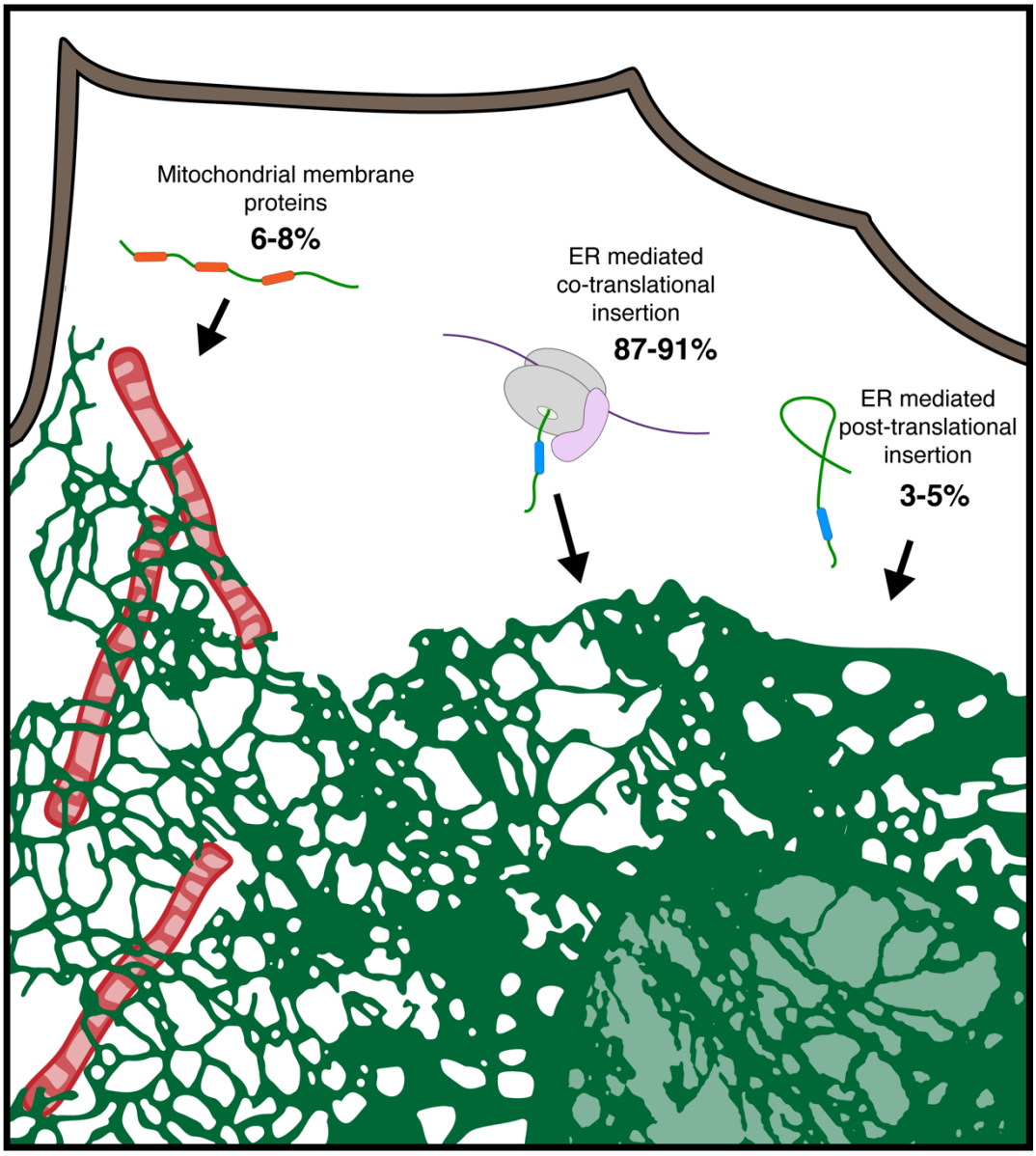


Figure 1.1 – The logistics of making the membrane proteome. The diagram depicts the three major types of membrane proteins as classified by their site of initial insertion into the lipid bilayer. The primary pathway for membrane protein biogenesis is the co-translational pathway, inserting proteins into the membrane of the endoplasmic reticulum [ER (shown in green), the nucleus is shown in light green]. Two other pathways facilitate the insertion of proteins that either evade the co-translational pathway or must be inserted into a different membrane system entirely [mitochondrial membrane proteins (MMPs)]. The majority of MMPs are inserted directly into the mitochondrial membranes after their synthesis by cytosolic ribosomes. The membrane embedded regions are kept insertion competent by cytosolic chaperones until their arrival and integration by specialized mitochondrial translocases. Mitochondria are depicted in red and their membrane embedded regions (orange) are depicted as distinct from ER destined TMDs (blue). Tail-anchored proteins destined for the ER are depicted on the right and have dedicated cytosolic and membrane components to keep them insertion competent and facilitate their insertion post-translationally, respectively. Plasma membrane is shown in dark grey.

Co-translational recognition and targeting of IMPs

Prior to the identification of machineries involved in membrane protein insertion, early biochemical studies had both predicted (Milstein et al., 1972) and then established the minimal requirements for secretory protein translocation: (1) a signal encoded in the protein sequence (Blobel and Dobberstein, 1975), (2) a cytosolic factor recognizing the targeting signal (Walter and Blobel, 1980), and (3) a membrane resident factor facilitating translocation (Meyer and Dobberstein, 1980; Walter et al., 1979). Following these initial observations, later studies narrowed the minimal requirements for targeting to a stretch of 7-15 hydrophobic residues that can adopt an α-helical secondary structure (Rothe and Lehle, 1998; Valent et al., 1995). Furthermore, the signal recognition particle (SRP) was unambiguously defined as the primary factor involved in the recognition and targeting of signal sequences (Walter and Blobel, 1982). Composed of multiple proteins and an RNA scaffold, the SRP is ideal for complementary binding of a translating ribosome and an ER targeting element. Indeed, structural analysis of SRP associated ribosomes revealed tight interactions between the two that place the functional domain of SRP (SRP54) near the ribosomal exit tunnel (Halic et al., 2004; Schaffitzel et al., 2006; Voorhees and Hegde, 2015). Also called the M-domain, SRP54 is unusually enriched with methionine residues which line the hydrophobic groove responsible for signal engagement (Keenan et al., 1998; Zopf et al., 1990). Both the moderate hydrophobicity and flexibility of methionine residues provides a molecular explanation for the broad promiscuity of SRP and its role as the primary targeting factor. The general principles established by studying signal sequences nicely explained why TMDs, which are usually longer and more hydrophobic, are also recognized and targeted to the ER co-translationally (Ulbrandt et al., 1997). Modern genome-wide studies have confirmed SRP mediated targeting of nearly all ER destined TMDs, and the majority of signal-peptide containing proteins (Costa et al., 2018; Schibich et al., 2016).

Two remarkable features of the SRP-mediated targeting pathway are the selectivity of SRP and the incredible efficiency by which substrates are targeted to the ER membrane. While the molecular mechanism remains ambiguous, SRP specificity can be attributed to its ability to recognize ribosomes translating hydrophobic stretches, even before a hydrophobic element emerges from the exit tunnel (Chartron

et al., 2016; Ogg and Walter, 1995; Voorhees and Hegde, 2015). This allows SRP to "wait" for a signal to emerge and engage it immediately. Complementing this pre-engagement strategy, targeting to the ER relies on the unique energy dependent interaction between SRP and its cognate receptor, the SRP receptor (SR) (Gilmore et al., 1982a, 1982b). The SRP and SR both contain guanosine triphosphate (GTP) hydrolase (GTPase) domains, but no auxiliary nucleotide exchange factor or GTPase activating protein. Instead, GTP hydrolysis is induced via the SRP-SR interaction at the ER and substrate release requires GTP hydrolysis (Connolly and Gilmore, 1989; Powers and Walter, 1995). Thus, a hydrophobic element remains shielded in the cytosol and cooperative GTP hydrolysis ensures release only after successful targeting. The coincidence detection of a ribosome and ER targeting element, along with the unique reciprocal GTPase cycle, provides an extremely efficient targeting apparatus. These general concepts have been supplemented with numerous technical studies that nicely validate the SRP-SR pathway as both necessary and sufficient for co-translational targeting of nearly all ER destined IMPs (Akopian et al., 2013; Keenan et al., 2001).

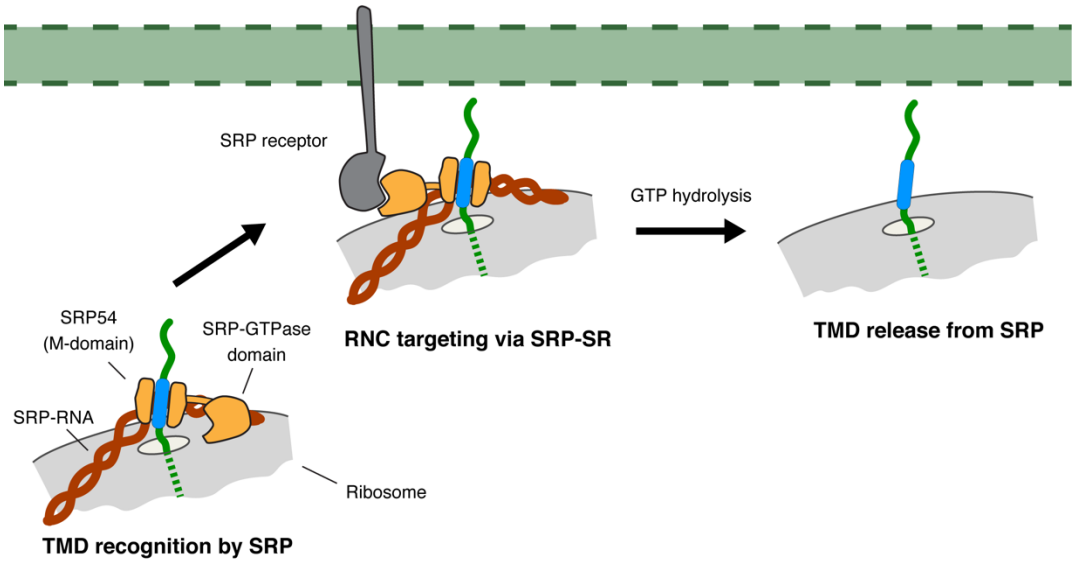


Figure 1.2 – SRP mediated targeting to the endoplasmic reticulum. Model depicting the key aspects of SRP mediated recognition and targeting in the cytosol. SRP is positioned on the ribosome through interaction with its scaffold RNA and various protein elements interacting near the exit tunnel. As a hydrophobic element emerges from the ribosome exit tunnel, SRP54 (M-domain) is positioned to recognize and capture this TMD via hydrophobic elements. Targeting of the entire ribosome nascent chain complex (RNC) is facilitated by binding of the SRP GTPase domain to the GTPase domain of SRP receptor at the ER membrane. Cooperative GTP

hydrolysis causes release of the TMD so that it is free to engage insertion machinery present in the ER membrane (depicted in green).

The translocation associated machinery

After targeting of a ribosome nascent chain complex (RNC) to the surface of the ER membrane, a TMD is now in the correct post code. Long before any mechanistic understanding of membrane protein insertion, it was postulated that a proteinaceous pore embedded within the ER membrane anchored the ribosome, facilitated insertion of TMDs, and translocated polypeptide across the membrane (Blobel, 1980). This hypothesis was consistent with a handful of initial observations. First, membrane bound ribosomes were observed tightly bound to isolated ER microsomes (Adelman et al., 1973). Second, puromycin induced release of ribosome peptide products were preferentially translocated into the lumen of ER vesicles (Redman and Sabatini, 1966). Finally, the observed ribosome-membrane interaction was prevented when ER microsomes were subjected to limited proteolysis (Borgese et al., 1974). However, a lack of easily manipulated functional assays hampered the identification of this putative "translocon". Fortunately, the principles established from the SRP targeting pathway provided a framework for the development of genetic and biochemical tools that ultimately led to its discovery.

Leveraging the sufficiency of signal sequences in targeting soluble proteins to the ER lumen, Deshaies and Schekman designed a genetic screen in yeast with an essential cytosolic enzyme artificially fused to a signal sequence (Deshaies and Schekman, 1987). Sequestration in the ER lumen was lethal and any mutants restoring growth were indicative of failed translocation. Subsequent characterization revealed mutations in three essential genes termed Sec61p, Sec62p, and Sec63p (Deshaies and Schekman, 1987; Rothblatt et al., 1989). Sec61/62/63 could be purified as a complex that was embedded in the ER membrane and therefore was an ideal candidate (Deshaies et al., 1991). Further genetic and biochemical studies established Sec61p as the core component necessary for both secretory protein translocation and membrane protein integration (Sanders et al., 1992; Stirling et al., 1992). The mammalian homolog of Sec61p is tightly associated with ribosomes and directly interacts with translocating nascent chains (Görlich et al., 1992a; Müsch et

al., 1992; Sanders et al., 1992). Together, these observations led to the purification of the heterotrimeric Sec61complex consisting of Sec61α, β, and γ (referred to as Sec61). When reconstituted into purified proteoliposomes, mammalian Sec61 was indeed confirmed to be the minimal machinery for translocation and insertion of model secretory and IMP substrates (Görlich and Rapoport, 1993) (Fig. 1.3).

A clear model for secretory and membrane protein biogenesis had begun to emerge with SRP targeting all hydrophobic elements, and Sec61 either integrating TMDs or translocating soluble polypeptide across the membrane (Fig. 1.3). However, both SRP-independent targeting and Sec62/63-dependent translocation had been observed. These conflicting results were reconciled with the discovery of a post-translational pathway. Substrates containing relatively hydrophilic signal sequences that escape recognition by SRP are kept translocation competent by cytosolic chaperones and delivered to the ER resident Sec62/63 complex (Plath and Rapoport, 2000). Sec62/63 partners with Sec61 and the luminal ATPase BiP in order to ratchet the soluble protein across the ER membrane (Matlack et al., 1999; Panzner et al., 1995; Plath et al., 1998). Two important points are worth highlighting. First, Sec62/63 occupies the ribosome binding site on the cytosolic face of Sec61α and therefore is mutually exclusive with the co-translational pathway (Itskanov and Park, 2019; Wu et al., 2019). Second, the post-translational system is specific for some secretory proteins and no IMPs appear to use this pathway (Costa et al., 2018). Thus, Sec61 is a generalist at the center of translocation but has limitations that can be overcome with aid from trans-acting factors to accommodate problematic substrates.

Many other translocation associated proteins have been identified and characterized. Factors such as the oligosaccharyl transferase (OST) complex, signal peptidase, and ER resident chaperones such as calnexin (CNX) play an essential role in protein folding and biogenesis. Additionally, several factors have been demonstrated to be in close association with Sec61 and engage a signal peptide or TMD during translocation. Examples include the translocating-chain associating membrane protein (TRAM) and the signal sequence receptor alpha (later termed TRAPα) (Görlich et al., 1992b; Wiedmann et al., 1988). However, the molecular contribution of these associated factors remains to be determined. Nevertheless, these studies illustrate

Sec61 is the primary conduit into the ER membrane but nearby auxiliary factors can also act post- or co-translationally on an incoming substrate.

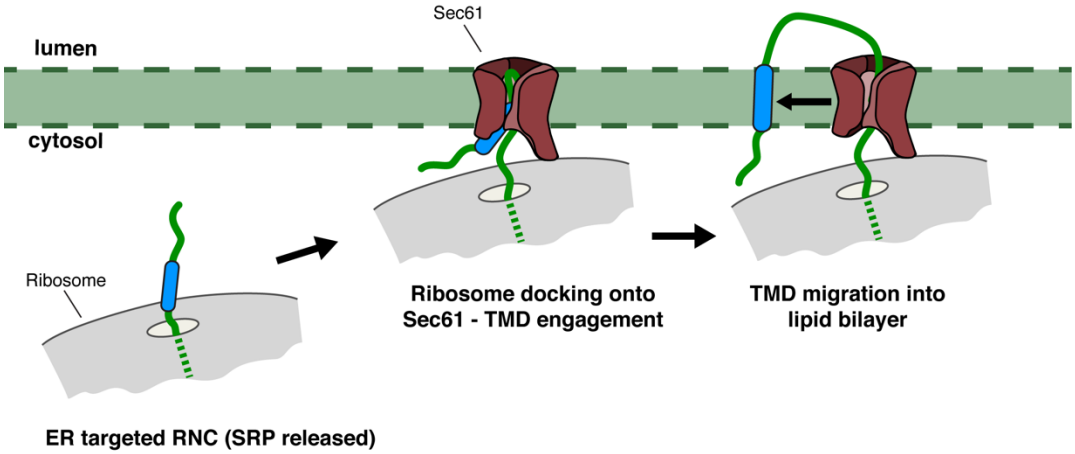


Figure 1.3 – A model for Sec61 mediated insertion of TMDs. Based on early biochemical and genetic studies, this model prevailed as the mode of TMD insertion. A newly targeted RNC would engage Sec61 via a high affinity interaction between the ribosome and Sec61. This places the TMD to engage Sec61 and then follow the most energetically favorable path into the membrane. Sec61 provides both a lateral path into the membrane for TMDs and a path through the membrane for soluble polypeptide being translocated into the ER lumen. For example, secretory proteins or soluble domains of membrane anchored proteins.

The complexities of membrane protein biogenesis

While the exact function of many translocation associated proteins remains unknown, a role beyond insertion has been established for several trans-acting factors. There is a critical time window after targeting and during insertion where several post-translational/post-translocational modifications will occur that can dramatically affect the maturation of nascent membrane proteins. For example, cleavage of a signal peptide occurs shortly after the signal leaves the translocon and is recognized by the translocon associated signal peptidase I. This specialized serine protease contains a hydrophobic groove to recognize the signal and place its C-terminal region in the active site for proteolytic cleavage (Paetzel et al., 1998). Failure to remove a signal from newly translocated proteins can cause misfolding and loss of function, explaining several human diseases resulting from inefficient cleavage (Ito et al., 1993; Racchi et al., 1993).

For a large subset of membrane proteins, even modest modifications can affect their folding, maturation and trafficking. A primary influencer at this early stage

is N-linked glycosylation. Nearly immediately after a nascent chain emerges from Sec61 it is scanned by the closely apposed OST for a glycosylation motif (Asn-X-Thr/Ser), where a core oligosaccharide can be covalently attached to the side chain nitrogen of Asn (Helenius and Aebi, 2004; Nilsson and Von Heijne, 1993). The degree to which glycans affect membrane protein biogenesis is highly variable but direct and indirect influences have been reported. The modest contribution of glycans to protein folding is because, apart from the covalent linkage, there are few direct contacts between sugar moieties and the protein surface (Wormald et al., 2002). Instead, a glycan most likely influences protein folding by biophysically constraining the conformational profile during the initial stages of protein folding, as seen with short polypeptides (Imperiali and O'Connor, 1999). This hypothesis nicely explains several observations: 1) N-glycosidic linkages are commonly observed adjacent to secondary structure transitions (Petrescu et al., 2004), 2) already folded proteins are not destabilized by enzymatic removal or modification of glycans (Helenius and Aebi, 2004), and 3) glycans can directly increase the stability, solubility, and protease resistance of at least some proteins (Imberty and Pérez, 1995; Imperiali and O'Connor, 1999; Kundra and Kornfeld, 1999; Wormald and Dwek, 1999). Additionally, the importance of glycans is reflected in the partial reduction in folding or increased temperature sensitivity observed upon their removal (Helenius, 1994; Olden et al., 1982). An important aspect of glycobiology is that glycans remain somewhat independent appendices. This is important because it provides a means to bind a nascent polypeptide and monitor the folding status without perturbing protein structure or folding. The best understood example of this is the lectins CNX and calreticulin (CRT), which can recognize incompletely folded proteins through the presence of a single glucose (added to the core glycan structure), and upon their recruitment they act as molecular chaperones to prevent aggregation and subsequent degradation (Helenius and Aebi, 2004). Removal of this glucose by specific glucosidases signals protein folding, disrupts CNX and CRT binding, and permits trafficking out of the ER. Conversely, prolonged association will eventually direct the protein for degradation via ER associated degradation (ERAD) machineries (Trombetta, 2003). Thus, the addition of glycans to a protein can affect protein folding locally by providing biophysical constraints but can also influence folding and

maturation on a more global scale through recruitment of various chaperones and quality control factors.

An even more influential modification on membrane protein biogenesis is the formation of intramolecular disulfide bonds. Not only do certain membrane proteins strictly require them for folding (Perlman et al., 1995), but this greatly expands the conformational and functional scope of the mammalian proteome (Buchanan and Balmer, 2005), explaining why the number of disulfide bonds is greater in higher organisms (Bošnjak et al., 2014). In mammalian cells the large majority of disulfide bonds are formed in the oxidizing environment of the ER lumen, during the initial stages of protein biogenesis (Sevier and Kaiser, 2002). Their formation is catalyzed by a family of chaperones known as thiol-disulfide oxidoreductases. As a general rule this family of enzymes mediates thiol-disulfide exchange reactions, in which the enzyme active site contains a disulfide bond that is reduced in order to oxidize the free thiols in its substrate (Noiva, 1999). Particularly relevant to membrane protein folding is protein disulfide isomerase (PDI), which acts both as a protein chaperone and is impressively versatile in its ability to oxidize, reduce, or isomerize disulfide bonds (Noiva, 1999). Effectively, this allows PDI to not only aid in the formation of disulfides but also monitor and isomerize incorrectly paired cysteines. Thus, disulfide bond formation is a critical aspect of membrane protein folding and as such is carefully monitored to improve the fidelity and efficiency of biogenesis.

There are two important points to keep in mind in regard to the relationship between membrane protein insertion/folding and the nuanced complexities of biogenesis. The first is that while many of these processes, such as glycosylation and disulfide bond formation, can occur concomitantly with co-translational insertion and folding, they are not strictly required. For example, many proteins can fold efficiently without their native glycans and a large number of proteins do not contain disulfide linkages (Bošnjak et al., 2014; Helenius and Aebi, 2004). Thus, these auxiliary modifications and the factors responsible can act in concert with insertion and folding machineries but are not obligately coupled. Secondly, an extremely important role for post-translational modifications is they provide a direct link to quality control mechanisms such as ERAD. Nascent membrane proteins that are either incompletely or incorrectly folded will have a longer resident time on factors such

as calnexin/calreticulin [untrimmed glycans, (Helenius et al., 1997)], PDI [oxidized thiols, (Tsai et al., 2001; Wang and Chang, 2003)], and general chaperones such as BiP [exposed hydrophobic patches, (Blond-Elguindi et al., 1993)]. As a result, specialized machineries will recognize these problematic substrates and target them for degradation (Reddy and Corley, 1998). This relationship necessitates a homeostatic balance between membrane protein biogenesis and degradation because the modes of substrate recognition are closely linked. In other words, if either process is left unchecked it can saturate or "out compete" the other, a very important consideration when monitoring the biogenesis of membrane proteins in cells or in vivo.

Sec61 and the mechanism of TMD insertion

The pathway into the membrane via Sec61 became clear upon solving the structure of the archaeal homolog (van den Berg et al., 2004) (Figure 1.4). Sec61α is the functional unit consisting of 10 TMDs forming an hourglass shaped channel with a luminal plug to maintain the permeability barrier. Biophysical, biochemical and structural studies support plug displacement upon engagement of an RNC, thus opening a hydrophilic pore that is semi-continuous with the ribosomal exit tunnel (Beckmann et al., 1997; Crowley et al., 1993, 1994; Simon and Blobel, 1991). The membrane embedded architecture of Sec61α is arranged pseudo-symmetrically to form a vice-like structure with the β and γ subunits flanking the rear hinge region and a lateral gate between TMDs 2 and 7 (Plath et al., 1998). The lateral gate seems to passively open upon engagement of a suitably hydrophobic signal, explaining the observed (near-simultaneous) interactions of a TMD with both lipid and Sec61 (Do et al., 1996; Heinrich et al., 2000; Martoglio et al., 1995; Mothes et al., 1997). Structural analysis of ribosome bound Sec61 complexes during translocation confirmed these prior biochemical observations and provide a clear 2-step model for translocation (Voorhees and Hegde, 2016; Voorhees et al., 2014). Upon ribosome binding at the cytosolic face, Sec61 enters a "primed" state where the lateral gate becomes slightly cracked in a conformation that is more amenable to binding of a hydrophobic signal. However, this primed state does not alter the integrity of the luminal plug or open the channel. Engagement of a signal displaces all interactions between

TMDs 2 and 7, as well as the luminal plug, and opens the channel both laterally and axially (Voorhees and Hegde, 2016) (Figure 1.4).

One important point worth highlighting is that Sec61 is mechanistically constrained, despite its ability to service a large range of substrates. For example, priming of the channel upon ribosome binding exposes a hydrophobic patch that must be displaced (Voorhees and Hegde, 2016). Presumably, this serves to triage signals for the minimal hydrophobicity requirements. Consistent with this notion, hydrophilic signal peptides fail to be translocated even after productive targeting by SRP (Jungnickel and Rapoport, 1995). Furthermore, the hydrophobicity threshold is bypassed by mutations disrupting this patch or the integrity of the closed channel (Derman et al., 1993; Junne et al., 2007; Osborne and Silhavy, 1993; Smith et al., 2005; Trueman et al., 2012). Thus, signal engagement and opening of the channel is the rate limiting step. Prior to this, no polypeptide can enter the channel because the pore remains completely closed, even in the primed (ribosome bound) state (Voorhees et al., 2014). With this concept in mind, it is not surprising that TMDs challenged by poor hydrophobicity and unusual length are inefficiently inserted by Sec61 (Hessa et al., 2005, 2007; Jaud et al., 2009). These limitations become incredibly important when considering the insertion of multipass (multispanning/polytopic) membrane proteins, which are enriched with biophysically challenging TMDs (Baker et al., 2017).

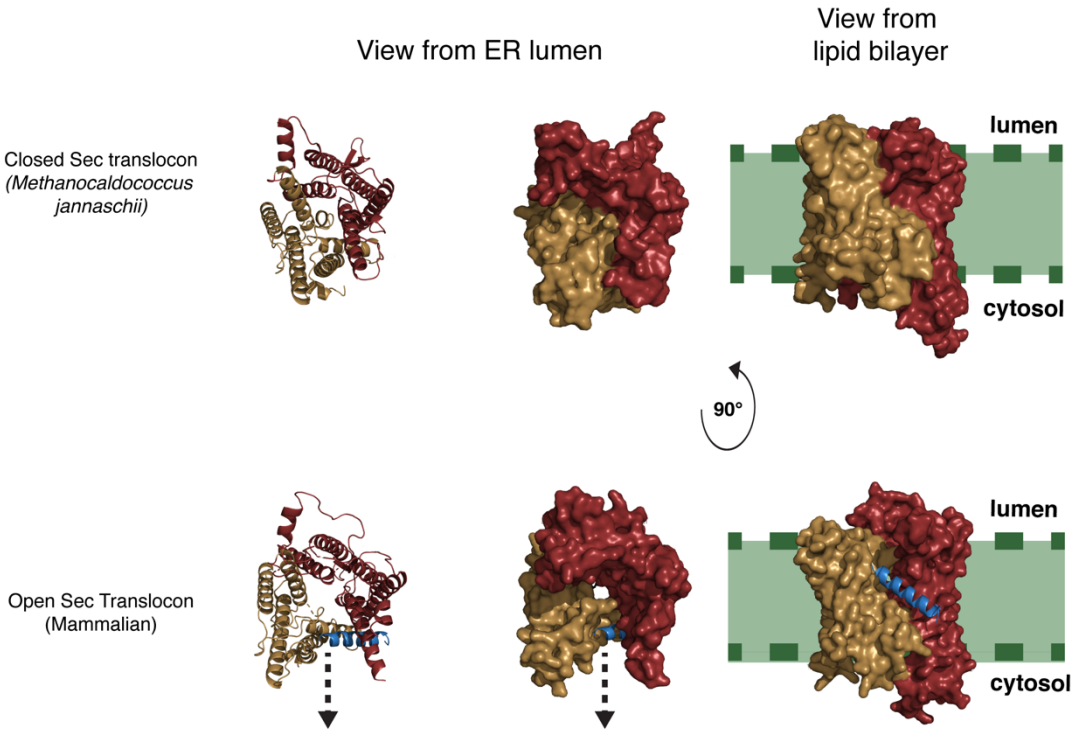


Figure 1.4 – The structural mechanics of Sec61 mediated insertion. The figure depicts two structures of the highly conserved Sec6α subunit of the Sec61 complex. The left most panels are cartoon representations of the peptide backbone, the center and right panels are surface representations. The two pseudo-symmetric halves of Sec61α are colored in tan (TM1-5) and maroon (TM6-10). The top panels represent a structure of Sec6 from an archaeal homolog in a completely closed state (Van den Berg et al., 2004). The bottom panels represent the structure of mammalian Sec61 bound to a ribosome (not shown) and engaged by a signal peptide (blue cartoon) (Voorhees and Hegde, 2016). Notice, when engaged by a hydrophobic element (in this case a signal peptide) two major conformation changes occur. First, the luminal plug is displaced and a pore through the membrane opens to allow polypeptide translocation. Second, interactions between TMD2-7 at the lateral gate are broken and a pathway into the lipid bilayer becomes available. The signal peptide is places right at the lateral gate making key contacts with TMD2 of Sec61α, displacing it open the lateral gate. Comparing the views from the lipid bilayer, one can appreciate how the hydrophobic α-helix is positioned well to migrate into the lipid bilayer.

Insertion of multipass membrane proteins

A mechanistic understanding of the targeting and translocation machinery came from a library of studies examining and manipulating simple model substrates. This included robust secretory proteins and idealized TMDs that are ~20 residues in length and particularly hydrophobic. However, the majority of TMDs are not simply membrane anchors but fundamental to protein function. The conserved sequence features that confer function often place major biophysical constraints on TMD insertion and IMP folding. The simplest example is multipass membrane proteins,

where intramembrane interactions between TMDs results in a unique 3-dimensional fold (Illergård et al., 2011). This necessarily requires charged and polar residues within the TMD, explaining their high frequency in multipass IMPs (Baker et al., 2017). Another biophysical constraint is a wide variation in both length and sequence characteristics due to differences in lipid composition at their final location (Sharpe et al., 2010), as well as unique folds which force TMDs away from a perpendicular arrangement in the lipid bilayer (Yernool et al., 2004). Furthermore, multipass membrane proteins must be oriented correctly. While single pass IMP topology is heavily biased by the "positive-inside rule", with charges flanking their single TMD (Hartmann et al., 1989; von Heijne, 2006), bioinformatic analysis of multipass TMDs revealed significantly less charge bias (Baker et al., 2017). Nevertheless, they all must be inserted into the ~30 angstrom lipid bilayer of the endoplasmic reticulum. The competing constraints of a single co-translational integration machinery and an extremely diverse set of TMDs imparts a massive challenge on multipass IMP biogenesis

How the cell overcomes this challenge remains entirely unclear. It has long been postulated that additional factors may act in concert with the translocation machinery in order to accommodate the large substrate pool (Shao and Hegde, 2011). In this thesis, I will discuss two of our recent contributions that shed light on the process of multipass IMP biogenesis. The first study characterizes a new molecular player in membrane protein insertion and topogenesis, while the second study identifies and characterizes a novel intramembrane chaperone required for multipass IMP biogenesis and folding.

Chapter II

EMC is required to initiate accurate membrane protein topogenesis

The topogenesis problem

All TMDs within a multipass IMP are thought to have defined and fixed orientations relative to the bilayer, allowing flanking soluble domains to be segregated amongst the two sides of the membrane (Blobel, 1980; Katz et al., 1977). Accurately establishing an IMPs topology is therefore essential for proper maturation and function. As an example, the structure of most G-protein coupled receptors (GPCRs) require exoplasmic disulfide bonds for productive folding (Perlman et al., 1995). Disulfide bond formation occurs during biogenesis and is stabilized by the oxidizing environment in the ER lumen, away from the reducing capacity of the cytosol. Consequently, insertion in the wrong orientation will result in misfolding and render it non-functional. Some proteins inserted incorrectly can even be dominantly detrimental (Hegde et al., 1998). Thus, accurate topogenesis is not only essential for membrane protein biogenesis but also to maintain overall protein homeostasis.

While the topology of a multipass IMP appears more complex, a simple examination of known protein structures reveals a strong topological interdependence between TMDs. As a general rule, each TMD will adopt the inverse orientation of the preceding TMD. Fixing the topology of the first TMD would therefore constrain the remaining TMDs and define the topological 'reading frame" (Blobel, 1980; Wessels and Spiess, 1988). This greatly simplifies the topogenesis problem for multipass IMPs, despite downstream TMD insertion complications (Öjemalm et al., 2012), and places topological control on recognition and insertion of the first TMD.

The most well-defined topological determinant is an N-terminal signal sequence and ~50-70 amino acids of a soluble domain preceding one or more TMDs. Co-translational engagement of Sec61 by the signal sequence initiates transport of the ensuing soluble domain into the ER lumen. This effectively commits the TMD to a topology where its flanking N- and C-terminus are exoplasmic and cytoplasmic, respectively (designated N_{exo}). Of the ~5000 predicted human membrane proteins inserted at the ER, ~31% contain N-terminal signal peptides and use this mechanism to establish topology (UniProt Consortium, 2018). Another ~5% are tail-anchored membrane proteins, whose topology is dictated by their post-translational insertion

(Hegde and Keenan, 2011). The remaining ~64% of membrane proteins are thought to rely on their first TMD for targeting and setting the protein's overall topology.

How the sequence information encoded within a signal anchor contributes to topology has remained fairly ambiguous. Several trends have emerged from mutagenesis studies demonstrating that length, hydrophobicity and flanking charges can influence a TMDs orientation (Higy et al., 2004). Additionally, for TMDs that can theoretically adopt either topology (short/flexible N-terminal domain), longer more hydrophobic TMDs and C-terminally flanking positive charges encourage the N_{exo} orientation (Kida et al., 2006; Wahlberg and Spiess, 1997). However, it has been extremely difficult to define specific rules as all these factors appear to play minor roles and many native signal anchors contain contradictory features (Higy et al., 2004; Ott and Lingappa, 2002). Only one absolute exists; an N-terminal domain that contains highly basic residues or folds prior to its translocation is retained in the cytosol (Beltzer et al., 1991; Denzer et al., 1995), and the signal anchor is forced into the respective topology (designated N_{cyt}). Despite these exhaustive studies on substrate contribution to topology, these trends are far from strict rules and little is known about how a cell correctly recognizes and orients signal anchors.

The primary question regards the mechanism by which these sequence features are decoded to determine signal anchor topology. Decades of work on the SRP-SR system established that all signal anchors arrive at the membrane via this pathway. The path of the TMD after SRP release, but prior to engagement of Sec61, is far less clear. The assumption that Sec61 is the next point of contact is based on two observations. Its high affinity for the ribosome, and end-point assays using crosslinking methods to show model signal anchors of either topology are in proximity to Sec61 (High et al., 1993a). Similarly, reconstitution studies demonstrated Sec61was sufficient for insertion of several model substrates, but necessity was only established for those containing a signal peptide (Görlich and Rapoport, 1993; Heinrich et al., 2000; Oliver et al., 1995). Collectively, a strong argument was made for Sec61 or its associated factors being the main contributors to topogenesis. However, an exhaustive analysis of structure informed Sec61 mutations revealed only modest defects in topology, even with a topologically challenged substrate (Goder et al., 2004; Junne et al., 2007).

Equally promising are the variety of proteins known to be near the site of translocation such as p180, RAMP4, the TRAP complex, TRAM, and others (Dudek et al., 2015). Of these, TRAM has been implicated in TMD insertion on the basis of its proximity and a stimulatory effect in proteoliposomes with purified Sec61 complex and SR (Do et al., 1996; Görlich and Rapoport, 1993; Heinrich et al., 2000). However, functional deficiencies in membrane protein topology or insertion have not been documented in loss of function studies. Furthermore, its closest homologs in yeast did not seem to have a role in translocation or TMD insertion (Barz and Walter, 1999), leaving in doubt any general and conserved role in membrane protein biogenesis.

More recently, the highly conserved **E**R **M**embrane protein **C**omplex (EMC) has been functionally and biochemically linked to membrane protein biogenesis. Since its discovery in yeast as a complex needed for ER protein homeostasis (Jonikas et al., 2009), EMC has been associated with highly pleiotropic phenotypes in many organisms (Bircham et al., 2011; Lahiri et al., 2014; Louie et al., 2012; Richard et al., 2013; Satoh et al., 2015; Savidis et al., 2016). Among them, several studies have documented reduced levels of various integral membrane proteins (Bircham et al., 2011; Richard et al., 2013; Satoh et al., 2015), some of which are retained in the early secretory pathway. The only process for which a direct role of EMC has been shown is the post-translational insertion of tail-anchored membrane proteins into the ER (Guna et al., 2017). This reaction can be reconstituted with purified EMC in liposomes, suggesting that EMC can directly facilitate TMD transfer from the cytosol into the lipid bilayer. Whether this biochemical activity is used for other types of membrane proteins that are inserted co-translationally remains a matter of speculation (Guna and Hegde, 2018).

The phenotypic consequence of EMC disruption for different types of multi-pass membrane proteins motivated us to investigate whether and how EMC might affect their biogenesis. Given earlier reports of EMC's effect on the levels of the GPCR rhodopsin in *Drosophila* (Satoh et al., 2015) and a related family member in yeast (Bircham et al., 2011), and the exceptional importance of GPCRs to nearly all aspects of human physiology (Hauser et al., 2017) we sought to understand the molecular basis of EMC's potential impact on this family of multi-pass membrane

proteins. Our biochemical dissection of GPCR biogenesis has revealed that EMC is a critical and previously unknown determinant of membrane protein topogenesis. EMC proved to be both necessary and sufficient for accurate insertion of many N_{exo} signal anchor sequences, a step critical to the biogenesis of GPCRs and many other membrane proteins.

EMC is required for optimal β_1-adrenergic receptor biogenesis in cells

Among the several membrane proteins reported to be impacted by EMC disruption, we chose GPCRs for several reasons. First, they are monomeric and relatively small (Rosenbaum et al., 2009), making them amenable to biochemical analysis in vitro (Sonnabend et al., 2017). This contrasts with very large (Louie et al., 2012) or multimeric complexes (Richard et al., 2013) whose reconstitution in vitro is daunting. Second, their 7-transmembrane topology is unambiguous and the structures of many examples are known (Venkatakrishnan et al., 2013). Third, the first TMD, which is a key determinant of topogenesis and serves as a signal anchor for ~90% of GPCRs, is highly diverse in sequence, properties, and flanking regions. This provides ample substrate diversity within a single class of proteins from which to investigate topogenesis. Finally, GPCRs are the largest class of multi-pass membrane proteins in vertebrates (~820 in human) and are of wide physiologic importance. Thus, GPCRs represented a tractable and important class of membrane proteins for investigating the potential role of EMC.

Earlier analysis had placed EMC's role at an early stage of rhodopsin functional expression in *Drosophila* (Satoh et al., 2015), but could not distinguish between effects on translation, maturation, degradation or trafficking. To investigate this, we analyzed post-translational effects of EMC disruption on the vertebrate β_1-adrenergic receptor (β_1AR) using a flow cytometry based assay. The C-terminus of a well-characterized β_1AR construct (Warne et al., 2008) was appended with GFP and RFP separated by a viral P2A sequence (Fig 2.1A). Translation of this mRNA will generate two products due to peptide bond skipping at the P2A sequence (De Felipe et al., 2006): the β_1AR-GFP fusion protein and a separate RFP. Thus, metabolically stable RFP serves as a 'counter' for the number of times this construct is translated, effectively integrating mRNA levels and translation efficiency into a single metric. Because one β_1AR-GFP is synthesized for each RFP, any reduction in GFP levels relative to RFP necessarily reflects post-translational degradation.

Relative to the baseline distribution of GFP:RFP ratios for the β_1AR reporter in wild type U2OS cells, the distribution was clearly reduced (by ~2-fold) in cells lacking EMC5, a core subunit of EMC (Fig 2.1B). Similar results were obtained in HEK293 cells disrupted for EMC6, a different core EMC subunit (Fig 2.1C), arguing against cell-type or subunit-specific effects. Acute reintroduction of EMC5 via an inducible promoter in EMC5-knockout cells restored the complete EMC (Guna et al., 2017) and completely rescued the reduced stability of the β_1AR reporter. Very similar effects of EMC disruption were observed for the tail-anchored protein squalene synthase (Fig 2.1B), a protein whose insertion into the ER is established to be EMC-mediated (Guna et al., 2017). Reporter cassettes lacking an insert or containing the single pass membrane protein, asialoglycoprotein receptor (ASGR1), showed no GFP:RFP ratio changes in EMC-knockout or rescue cells relative to wild type cells.

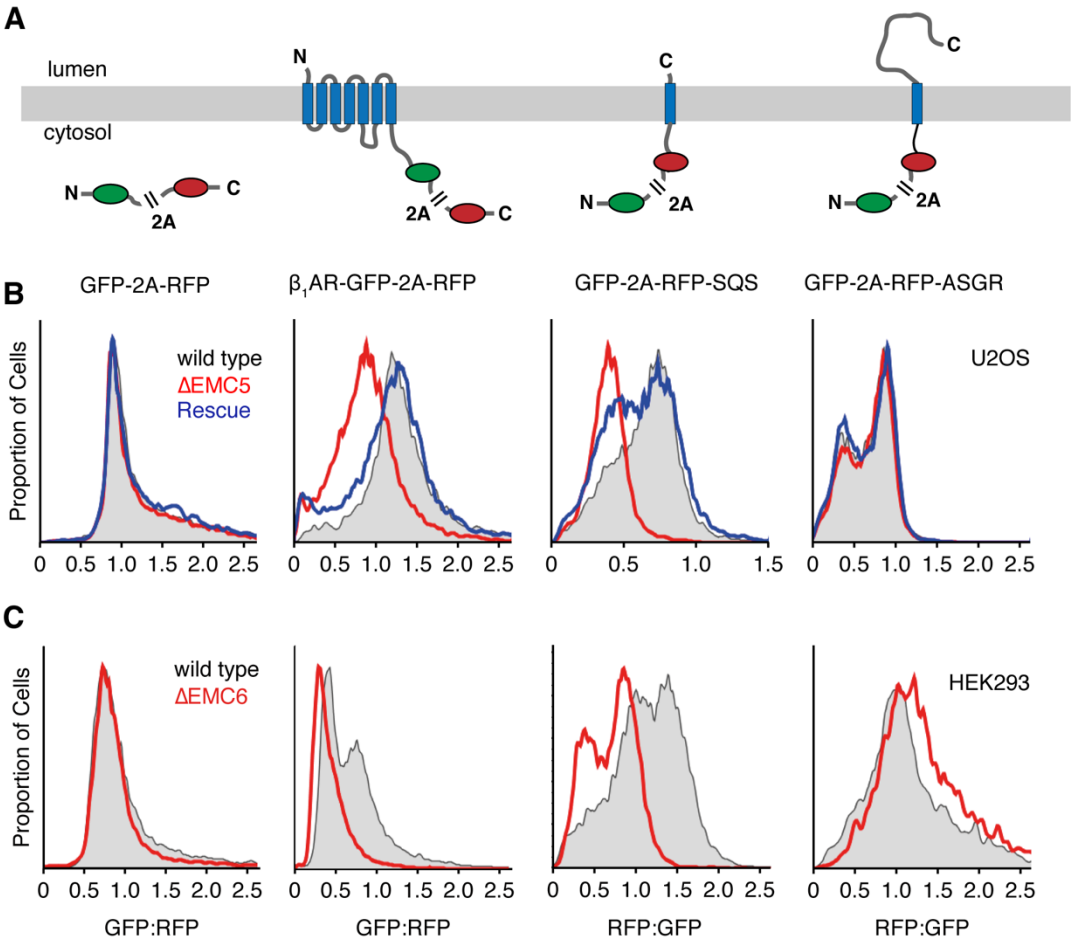


Figure 2.1 - EMC is required for optimal β1AR biogenesis in cells. (A) Diagram and topology of constructs for analysis of protein biogenesis by flow cytometry. All constructs contain GFP and RFP separated by a viral 2A peptide that mediates peptide bond skipping. Changes in the stability of a test protein fused to one of the fluorescent proteins changes the GFP:RFP fluorescence ratio. **(B)** Histograms of flow cytometry data monitoring the fluorescence protein ratio in the indicated U2OS cell lines for each construct. "ΔEMC5" indicates a knockout of EMC5, while "rescue" indicates ΔEMC5 cells rescued by inducible re-expression of a stably integrated EMC5. **(C)** Histograms of flow cytometry data monitoring the fluorescence protein ratio in unmodified (wild type) or EMC6-disrupted (ΔEMC6) HEK293 TREX cells. These experiments were done with help from S. Juszkiewicz.

These results indicate that post-translational β_1AR stability is dependent on EMC, the absence of which leads to its elevated degradation. The absence of any appreciable effect on ASGR1 excludes non-specific perturbation of protein biosynthesis or trafficking. More specifically, SRP-dependent targeting, Sec61-dependent insertion, and the stimulatory effect of TRAM, each of which is essential for optimal ASGR1 biogenesis (Görlich and Rapoport, 1993; Spiess and Lodish, 1986), are all apparently normal in EMC-disrupted cells.

Reconstitution of EMC-dependent β_1AR biogenesis in vitro

The altered β_1AR stability in EMC-disrupted cells is compatible with several explanations including altered biogenesis, trafficking, promiscuous degradation, and others. Of these, we favored an effect on biogenesis due to the previously observed EMC interaction with other biosynthetic factors such as OST and the Sec61 complex (Savidis et al., 2016). To investigate β_1AR biogenesis, we used an in vitro translation system composed of reticulocyte lysate (RRL) and ER-derived rough microsomes (RMs). This system recapitulates membrane protein insertion, but is not confounded by post-translational degradation or vesicular trafficking out of the ER.

In preliminary experiments, we established the conditions and assays to monitor membrane insertion, topology, and folding of newly synthesized ^{35}S-methionine-labeled β_1AR (Fig. 2.2). Correct topogenesis was inferred by a combination of

glycosylation, selective accessibility to cytosolically added protease, and immunoprecipitation (Fig. 2.3).

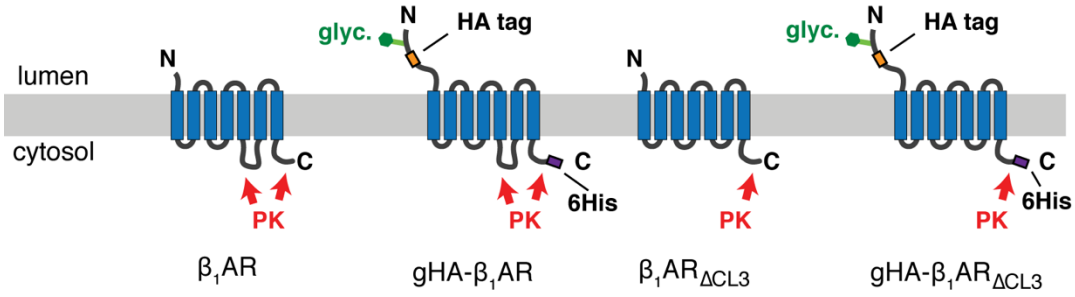


Figure 2.2 - System for assessing insertion and folding of β_1AR in-vitro: Diagram of constructs used to characterize β1AR topogenesis. ΔCL3 refers to the shortening of the cytosolic loop 3 between TMD5 and TMD6. The sites that should be accessible to cytosolically added proteinase K (PK) are indicated for each construct.

Analysis of parallel cassettes containing either the native sequence of β_1AR, or a version with an N-terminal affinity tag and a glycosylation site (Fig. 2.2), confirmed accurate insertion of the tagged β_1AR (Fig. 2.3A). Glycosylation serves as a fiducial marker for N-terminal translocation, while the epitope tag allows for topological analysis in protease protection assays. When translated in the absence of membranes, radiolabeled β_1AR migrated at ~45kDa by SDS-PAGE while the tagged version migrated slightly slower on the gel (Fig. 2.3A, lane and 7). Approximately 50-60% of the product was glycosylated when RMs were included in the reaction and the construct contained a glycosylation motif (lane 10). In all constructs tested, exposure to cytosolic protease produced a protected fragment that was dependent on the presence of membranes but readily digested upon the addition of detergent (Fig. 2.3A-B), suggesting integration into the lipid bilayer. Based on its size, this fragment agrees with insertion of the first 5 TMDs, suggesting the third cytosolic loop within β_1AR (CL3) is accessible to protease (see Fig. 2.2). Similar experiments in which the large protease accessible loop was removed ($\beta_1AR_{\Delta CL3}$) confirmed this assessment, and produced a fragment whose size corresponded to insertion of all 7 TMDs (Fig. 2.3B). Furthermore, the inserted population appears to be capable of efficient folding as judged by protease resistance even in the presence of detergent (Fig. 2.3B). Consistent with this interpretation, this population binds to immobilized alprenolol, a β_1AR antagonist, and selectively elutes with the agonist isoproterenol

(Fig. 2.3C). These results demonstrated that ~50-60% of β1AR could be inserted properly in this system.

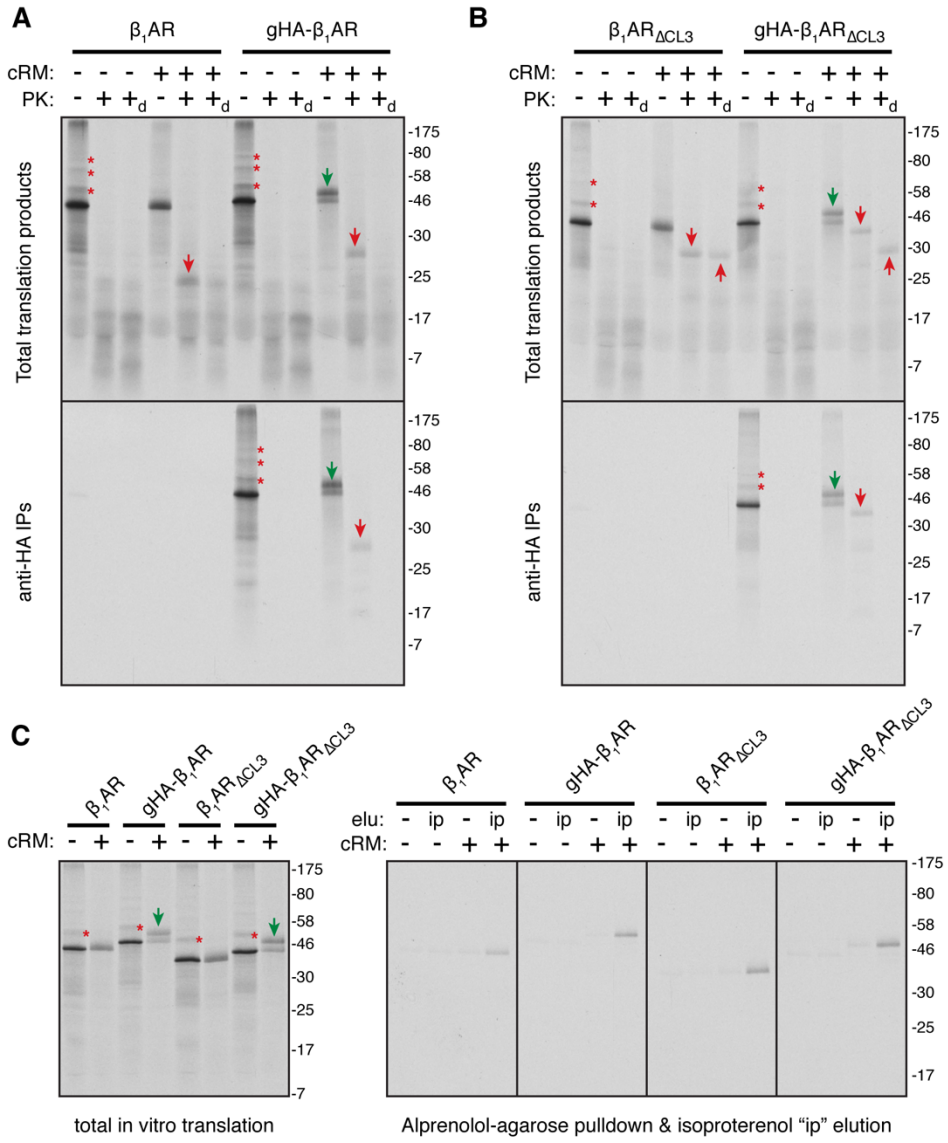


Figure 2.3 - Monitoring β1AR insertion in vitro. (A, B) ^{35}S-methionine labeled β1AR (or the N-terminally HA-tagged version with a glycosylation site) was translated in RRL in the absence or presence of canine pancreas-derived rough microsomes (cRM). The translation products were either left untreated or digested with proteinase K without or with detergent (subscripted d) as indicated. The samples were either analyzed directly (total translation products) or after immunoprecipitation via the N-terminal HA tag (anti-HA IPs) and analyzed by SDS-PAGE and autoradiography. Asterisks indicate ubiquitinated products; green arrows indicate glycosylated products; red downward arrows indicate PK-protected N-terminal fragments; red upward arrows indicate the protease-resistant 7-TMD core of β1AR left after digestion of the N- and C-terminal tails in the ΔCL3 variants. These assigned identities of the bands can be deduced by a combination of their size, change in migration upon addition of the N-terminal glycosylation site, change in digestion pattern upon shortening of CL3 to make

it proteaese-inaccessible, and IP via the HA epitope. **(C)** 35S-methionine labeled β1AR (or one of the indicated variants) was translated in RRL in the absence or presence of microsomes (cRM). An aliquot of the sample was analyzed directly (total in vitro translation) or solubilized and incubated with immobilized alprenolol (a β1AR antagonist). The resin was washed, then eluted in buffer without or with isoproterenol (ip; a β1AR agonist). Efficient recovery is only observed when β1AR is synthesized with cRM and eluted with isoproterenol. For all immunoprecipitations presented, the total represents 10% of the reaction and the IPs are 25% of the total reaction. These experiments were performed by S. Shao.

These assays provided three metrics which could be used to assess the consequences of EMC-disruption on β_1AR biogenesis using RMs derived from wild type (WT) versus EMC6-knockout (ΔEMC6) HEK293 cells: 1) protease protection, 2) glycosylation efficiency and 3) ligand binding. Importantly, accurate insertion of β_1AR was recapitulated in RMs derived from HEK293 cells, albeit with lower efficiency (Fig. 2.4-2.5). The protease-protected N-terminal fragment, diagnostic of correct topogenesis of the first five TMDs (Fig. 2.4A), was reduced by more than 50% in RMs from ΔEMC6 cells relative to WT cells (Fig. 2.4B). Furthermore, the deficiency was recapitulated in assays monitoring a glycosylated version of β_1AR, and a corresponding reduction in glycosylation efficiency was observed (Fig. 2.4C, top panel). Parallel experiments monitoring the ΔCL3 mutant revealed a similar disruption in integration of the stable 7TMD core (Fig. 2.4C, bottom panel), suggesting a rather dramatic instability during insertion and folding.

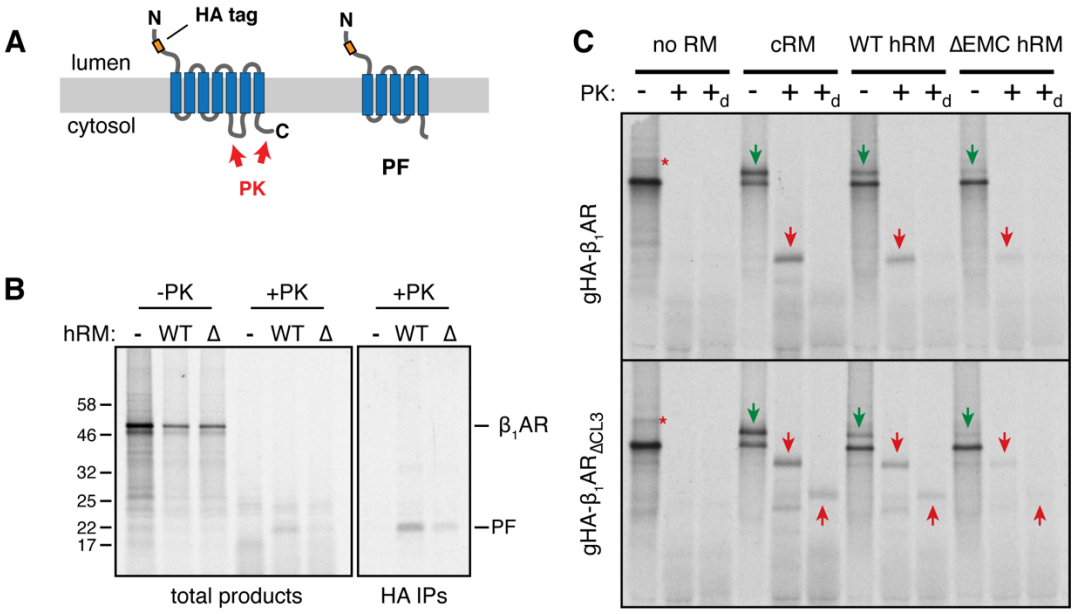


Figure 2.4 - Impaired insertion of β_1AR upon EMC knockout in vitro. (A) Diagram of the protease susceptible portions of an inserted β_1AR and the resulting protease resistant fragment (PF) indicative of correct insertion. **(B)** 35S-methionine labeled β1AR was translated in reticulocyte lysate (RRL) in the absence or presence of HEK293-derived rough microsomes (hRM) from wild type (WT) or ΔEMC6 (Δ) cells. The translation products were digested with proteinase K (+PK) or left untreated (-PK), then analyzed directly (total products) or after immunoprecipitation via the HA epitope tag (HA IPs). The positions of full length β1AR and the PF are indicated. **(C)** Protease protection assay on the indicated constructs performed as in A, but with either cRM or HEK293-derived microsomes (hRM) from either wild type (WT) or ΔEMC6 (ΔEMC) cells. Asterisks indicate ubiquitinated products; green arrows indicate glycosylated products; red downward arrows indicate PK-protected N-terminal fragments; red upward arrows indicate the protease-resistant 7-TMD core of β1AR left after digestion of the N- and C-terminal tails in the ΔCL3 variants. For all immunoprecipitations presented, the total represents 10% of the reaction and the IPs are 25% of the total reaction

Additionally, β_1AR in ΔEMC6 microsomes was less efficiently captured by immobilized alprenolol than in WT microsomes (Fig. 2.5A), while a folding-deficient construct (ΔTMD3) was not recovered at all. Further analysis in matched experiments revealed specific recovery of glycosylated β_1AR by alprenolol pull-downs (Fig. 2.5B), establishing glycosylation as an authentic read-out for inserted and folded β_1AR and further supporting EMC dependence. Despite these clear disruptions in biogenesis upon EMC deletion, equal amounts of β_1AR was recovered in

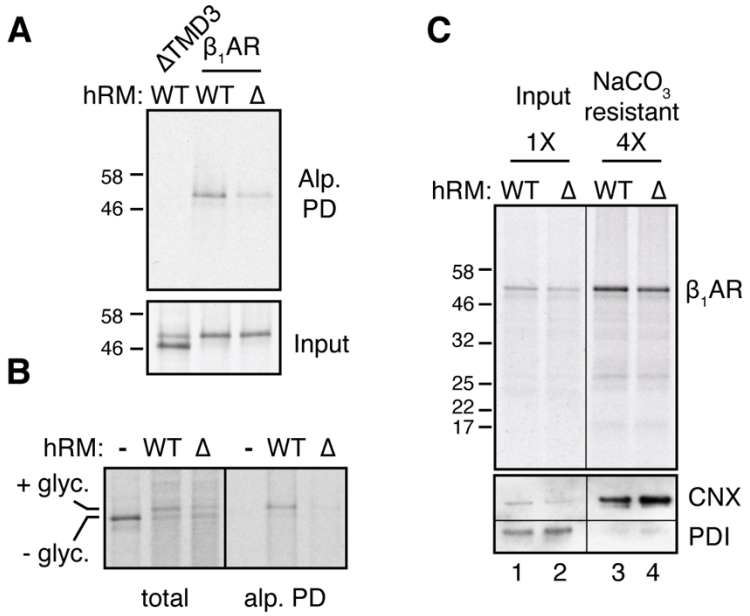


Figure 2.5 - EMC dependent folding of β_1AR. (A) ^{35}S-methionine labeled β1AR or a mutant lacking the third transmembrane domain (ΔTMD3) were tested for binding to immobilized alprenolol. The starting translation products (input) and alprenolol pulldowns (Alp. PD) are shown. **(B)** ^{35}S-methionine labeled gHA-β1AR was translated in reticulocyte lysate (RRL) in the presence of wild type (WT) or ΔEMC6 (Δ) hRM. The samples

were analyzed directly (total) or after selective capture by immobilized alprenolol and elution with isoproterenol (alp. PD). **(C)** ^{35}S-methionine labeled β1AR translation products produced in WT or ΔEMC6 hRMs were isolated by sedimentation of the hRMs and analyzed directly (input) or after extraction with Na2CO3 at pH 11.5 (Na2CO3 resistant; 4-fold excess was analyzed). β1AR was visualized by autoradiography, while the integral membrane ER protein calnexin (CNX) and ER-lumenal protein disulfide isomerase (PDI) were detected by immunoblotting. Unless otherwise indicated in the individual figure panel, all immunoprecipitations were loaded such that the total represents 10% of the reaction and the IPs are 25% of the total reaction

membranes pelleted from these two reactions (Fig. 2.5C, lanes 1, 2) and were similarly resistant to alkaline extraction (Fig. 2.5C, lanes 3, 4). Calnexin and protein disulfide isomerase (PDI) provided fiducial markers for the behavior of integral membrane proteins and soluble (extracted) luminal proteins, respectively. Although membrane targeting and integration of β_1AR is uninterrupted (i.e. equal alkaline resistance), there is a clear disruption in its biogenesis upon EMC depletion (glycosylation/ligand pull-downs).

Collectively, our prior characterization in-vitro allowed us to infer the fidelity of β_1AR biogenesis using a single metric, reduced glycosylation (Fig. 2.3-2.5). Glycosylation of the GPI-anchored prion protein (PrP) and the multipass membrane proteins TRAM2 were unaffected in ΔEMC6 microsomes (Fig. 2.6A-B). EMC independent insertion/translocation of these proteins was further supported by protease-protection assays. Furthermore, the biogenesis deficiency in ΔEMC6 RMs could not be overcome by using more microsomes in the reaction (Fig. 2.6C), suggesting an intrinsic problem in making β_1AR correctly rather than saturation in the in vitro system. Thus, the observed defect in β_1AR biogenesis cannot be explained by global defects in targeting, insertion or translocation.

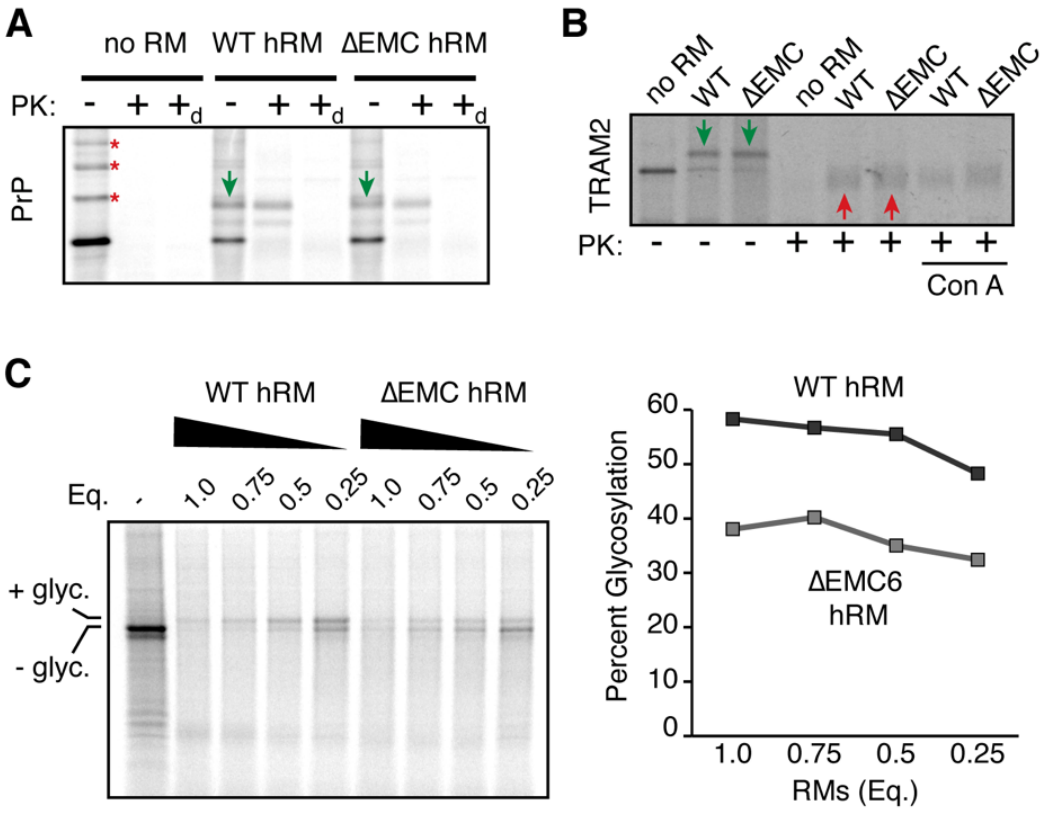


Figure 2.6 - An intrinsic β_1AR dependence on EMC. (A) ^{35}S-methionine labeled mammalian prion protein (PrP) was translated without or with the indicated hRM and analyzed by the PK-protection assay. Asterisks indicate ubiquitinated products; green arrows indicate doubly-glycosylated products (PrP contains two glycosylation sites). **(B)** ^{35}S-methionine labeled human TRAM2 was translated without or with the indicated hRM and analyzed by the PK-protection assay. Green arrows indicate TRAM2 glycosylated in the loop between the first and second TMD. After protease digestion, only the cytosolic-facing N- and C-terminal ends of the protein are digested, leaving behind a folded core (upward red arrows) comprising all eight TMDs. This product is recovered with ConA, verifying that it is the glycosylated central core. Note that no difference in TRAM2 glycosylation or protease-protection is seen between reactions performed with hRM from wild type or ΔEMC cells. **(C)** ^{35}S-methionine labeled gHA-β1AR was translated in RRL without or with wild type (WT) or ΔEMC6 (Δ) hRM at various relative concentrations (WT and ΔEMC hRM were normalized to have equal total protein concentration as judged by absorbance at 280 nm). The samples were analyzed directly (left panel) and the percent of translation product that is glycosylated was quantified by phosphorimager and plotted (right panel).

In direct comparisons using glycosylation or signal peptide cleavage we found that biogenesis of the secretory protein prolactin, the GPI-anchored prion protein PrP, and the membrane proteins ASGR1 and TRAM2 were unaffected in ΔEMC6 microsomes (Fig. 2.7A). These substrates represent the major types of model proteins analyzed in earlier work and collectively report on the integrity of SRP-dependent targeting, Sec61-mediated translocation and membrane insertion, the modulatory functions of TRAM and the TRAP complex, and the enzymatic activities of signal peptidase and OST. Indeed, immunoblotting verified that these components do not differ appreciably between WT and ΔEMC6 microsomes (Fig. 2.7B). Thus, the selective β_1AR biogenesis defect observed in EMC-deficient cells (Fig. 2.1) can be recapitulated in vitro (Fig. 2.7A). Furthermore, the observation that glycosylation of an acceptor site near the N-terminus of β_1AR (prior to the first TMD) is diminished in ΔEMC6 microsomes suggested that a relatively early step of β_1AR biogenesis may be impaired.

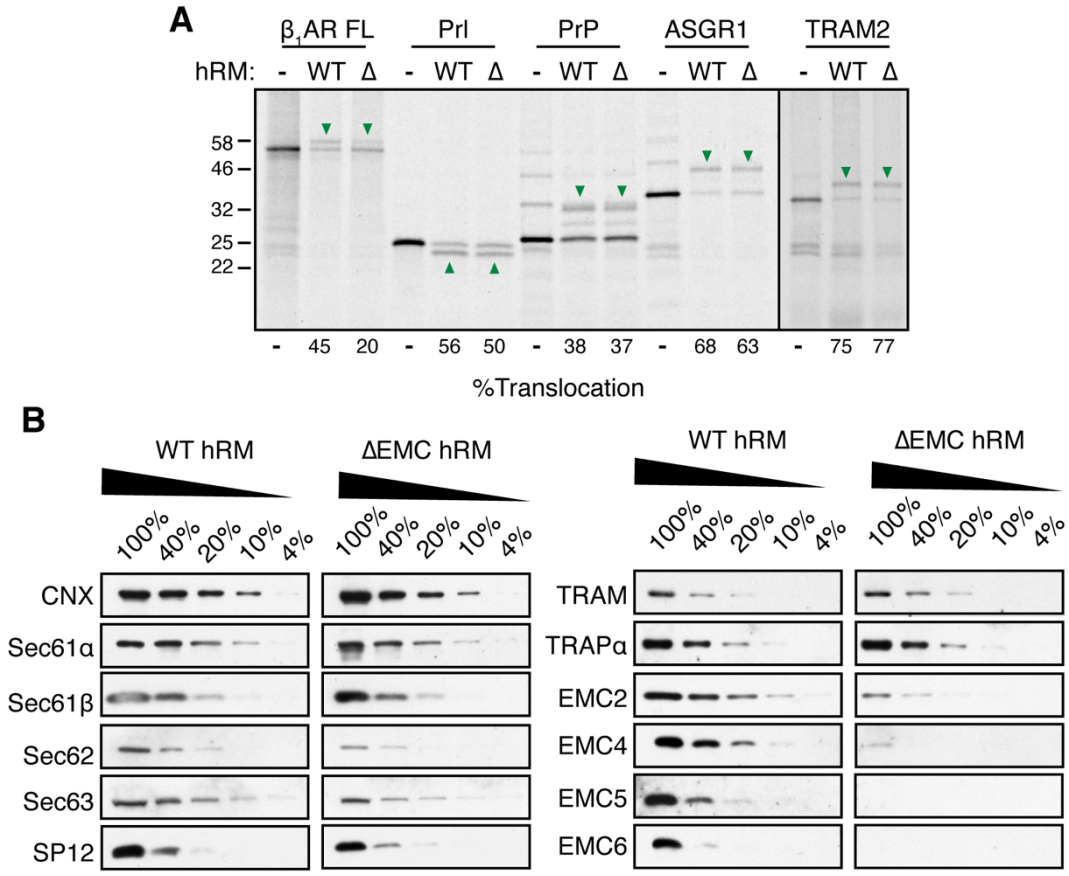


Figure 2.7 - A specific and direct role for EMC in β_1AR biogenesis. (A) The indicated proteins were translated without or with the indicated hRMs and analyzed for translocation by their glycosylation (downward green arrows) or signal peptide cleavage (upward green arrows). The % glycosylated or signal cleaved was quantified and shown below the gel. See also Fig. 2.6. **(B)** Different relative amounts of WT or ΔEMC6 hRMs were analyzed by immunoblotting for the indicated ER-resident proteins. Note that the WT and ΔEMC6 samples that are being compared were analyzed on the same gel and processed together.

EMC is required for accurate TMD1 topogenesis of β_1AR

To facilitate the analysis of early events in β_1AR biogenesis, we sought a simplified construct that still showed EMC-dependence. Serial truncations of β_1AR from the C-terminus revealed that the glycosylation defect was retained even in a construct that only contained the first TMD (Fig. 2.8).

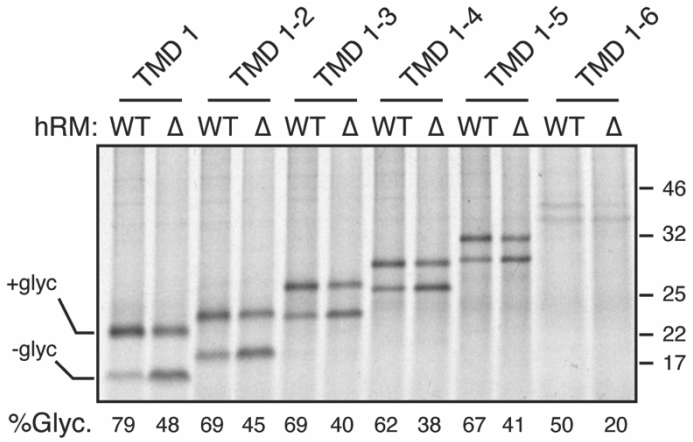


Figure 2.8 - EMC is required for biogenesis of a simplified β_1AR. (A) ^{35}S-methionine labeled gHA-β1AR constructs terminated after the indicated number of TMDs was translated in RRL in the presence of wild type (WT) or ΔEMC6 (Δ) hRM. The samples were analyzed directly (total) and the proportion of polypeptide that is glycosylated was quantified by phosphorimaging.

This simplified construct (termed β_1AR-TMD1) was effectively glycosylated at the N-terminus in WT microsomes, but significantly impaired in ΔEMC microsomes (Fig. 2.8, 2.9A). Protease digestion produced a protected fragment recovered by immunoprecipitation via a N-terminal HA epitope tag. As expected from the glycosylation analysis, this N-terminal fragment was substantially reduced in matched reactions containing ΔEMC6 microsomes (Fig. 2.9A). Thus, insertion of β_1AR-TMD1 in the correct topology is EMC-dependent, recapitulating the EMC-dependence of full length β_1AR in vivo and in vitro. Unexpectedly, pulldowns of the

same samples via a C-terminal His6-tag revealed a protease-protected fragment preferentially in the ΔEMC6 samples (Fig. 2.9A-B). This fragment was also seen at low levels in WT samples, albeit to a much lesser degree. Importantly, no protease protected fragments were observed in samples lacking RMs. This suggests that both WT and ΔEMC6 microsomes are comparably efficient in β_1AR-TMD1 insertion, consistent with the alkaline extraction results with full length β_1AR (Fig. 2.5C). However, the topology of around half of β_1AR-TMD1 molecules is inverted in ΔEMC6 microsomes, explaining the consistently impaired N-terminal glycosylation of truncated β_1AR constructs (Fig. 2.8). A β_1AR-TMD1 construct lacking the N-terminal glycosylation site also showed topologic inversion in ΔEMC6 microsomes (Fig. 2.9B), arguing against glycosylation influencing topogenesis.

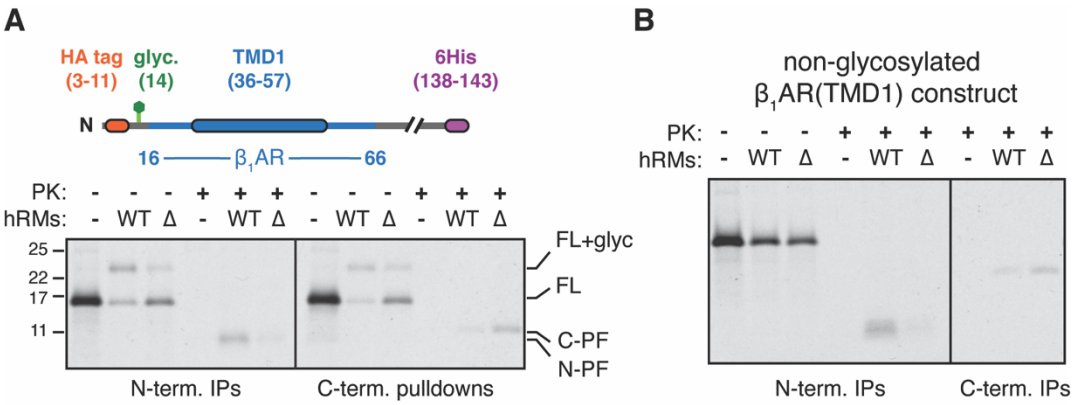


Figure 2.9 - EMC is required for accurate TMD1 topogenesis of β_1AR. (A, B) ^{35}S-methionine labeled β1AR-TMD1 (shown in the diagram) or a non-glycosylated version (B) was translated in the absence or presence of wild type (WT) or ΔEMC6 (Δ) hRMs, subjected to PK digestion as indicated, and the products recovered by either immunoprecipitation via the N-terminal HA tag (N-term. IPs) or pulldowns via the C-terminal His6 tag (C-term. pulldowns). The positions of unmodified full length (FL) product, glycosylated product (+glyc), and N- and C-terminal protease-protected fragments (N-PF and C-PF, respectively) are indicated. All immunoprecipitations were loaded such that the total represents 10% of the reaction and the IPs are 25% of the total reaction

Earlier analysis both in vivo and in vitro confirmed faithful targeting via the SRP pathway based on several well characterized substrates with known SRP dependence (Fig. 2.1, 2.6, 2.7). However, the insertion defect observed for a reporter consisting of only TMD1 suggested perturbations at very early steps, perhaps reflecting unanticipated disruptions in β_1AR TMD1 targeting. Direct analysis of SRP-

engagement was analyzed using ribosome-nascent chain complexes (RNCs) in which a ribosome fails to terminate translation of a truncated mRNA, producing a polypeptide tethered to the ribosome via the peptidyl-tRNA (Fig 2.10). Nascent chains of defined length were generated, placing the TMD outside the ribosomal exit tunnel near the site of SRP54 association. As expected for a signal anchor sequence (Chartron et al., 2016; Costa et al., 2018; Sakaguchi et al., 1987; Spiess and Lodish, 1986), 96-residue long cytosolic RNCs of β_1AR-TMD1 were associated with SRP similarly to the previously established N_{exo} and N_{cyt} model membrane proteins LepB (leader peptidase from *E. coli*) and ASGR1, respectively (Fig. 2.10).

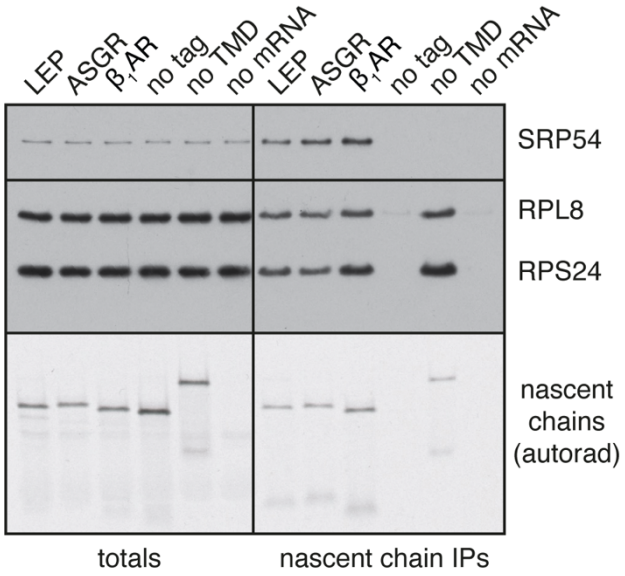


Figure 2.10 - β_1AR TMD1 faithfully engages SRP targeting pathway. ^{35}S-methionine labeled ribosome-nascent chains (stalled 39 residues downstream of the indicated TMDs) produced in reticulocyte lysate were affinity purified via an N-terminal FLAG epitope tag and analyzed by autoradiography to detect the nascent chains or immunoblotting for ribosomal proteins (RPL8 and RPS24) and SRP54. Controls either lacked a TMD or mRNA. All immunoprecipitations were loaded such that the total represents 10% of the reaction and the IPs are 25% of the total reaction. This experiment was performed by S. Juszkiewicz

To determine the point at which topogenesis diverges, we turned to the analysis of RNCs of different lengths representing intermediates in the targeting and insertion of β_1AR-TMD1 (Fig. 2.11A). When translated in the absence of membranes, protease digestion of such RNCs removes the exposed N-terminus, leaving behind a tRNA-associated C-terminal fragment protected by the ribosome [CTF2 (Fig. 2.11A-B)]. A minor, slightly larger product may either represent partial protection

by SRP, or some heterogeneity in the precise site of protease digestion. When 116-residue long β_1AR-TMD1 RNCs are presented to RMs, translocation of the N-terminus enables glycosylation, and this product is fully shielded from cytosolic protease by the ribosome and membrane [FL+glyc. (Fig. 2.11A-B)]. Relative to the situation in WT microsomes, ΔEMC6 microsomes show less glycosylation and less full-length protease protection (Fig. 2.11C). Instead, there is increased amounts of a non-glycosylated product whose N-terminus is accessible to protease (CTF1). Because the protected fragment is slightly larger than that seen in the absence of microsomes, it appears that the membrane affords protection of ribosome-proximal regions of the nascent chain (see Fig 2.11A diagram), reflecting insertion of the TMD in the N_{cyt} topology.

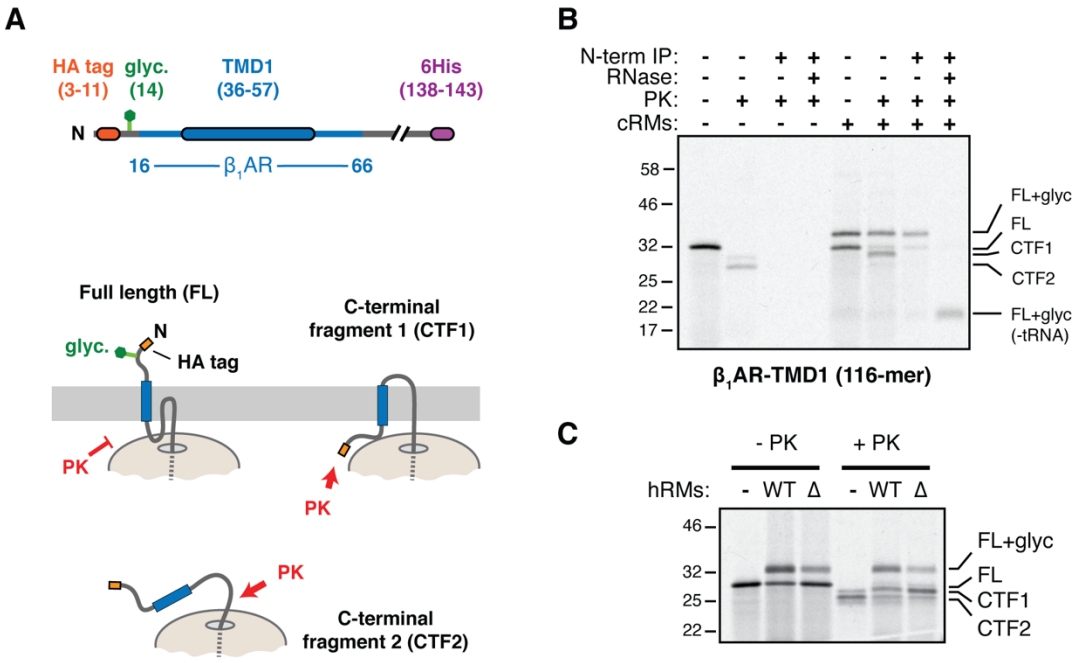


Figure 2.11 – A system to monitor early steps of β_1AR-TMD1 insertion. (A) Diagram representing the full length β_1AR-TMD1 and relative positions of the TMD, epitope tags and glycosylation site (top diagram). The bottom diagram shows the interpretation of the different products: Nexo-inserted nascent chains are glycosylated and fully protected from PK; non-inserted nascent chains are non-glycosylated and accessible to PK outside the ribosome and generate a C-terminal fragment (CTF2); Ncyt nascent chains are also non-glycosylated and accessible to PK, but have some regions protected by the membrane to generate a slightly larger C-terminal fragment (CTF1). **(B)** ^{35}S-methionine labeled 116-residue ribosome-nascent chain complexes of gHA-β1AR (see diagram) truncated 60 residues beyond the TMD were produced in RRL. They were incubated without or with canine pancreas-derived microsomes (cRMs) and subjected to digestion with PK as indicated. An aliquot of the PK-digested sample was subsequently immunoprecipitated via the N-terminal HA tag without or with RNAse

digestion as indicated. **(C)** As in B, 116-residue nascent chains of β1AR were targeted to WT or ΔEMC6 hRMs and analyzed by the PK protection assay. All immunoprecipitations were loaded such that the total represents 10% of the reaction and the IPs are 25% of the total reaction.

This difference in insertion between WT and ΔEMC6 microsomes is observed across a range of nascent chain lengths (Fig. 2.12). Of note, the difference was not as prominent for the 96-residue RNC that is truncated only 39 residues beyond the TMD (Fig. 2.12, right panel). At this length, the TMD has barely emerged from the ribosome and is just long enough for membrane insertion in the N_{exo} topology, but cannot achieve the N_{cyt} orientation. Thus, the EMC requirement can apparently be partially bypassed by constraining the RNC to only the N_{exo} option and providing it far more time for insertion than would be available during co-translational full-length β_1AR biogenesis. The fact that even this highly biased situation still shows an appreciable difference in insertion suggests that the deficiency observed in ΔEMC6 microsomes is not simply a kinetic problem; rather, the microsomes are intrinsically less capable of TMD1 insertion in the N_{exo} topology.

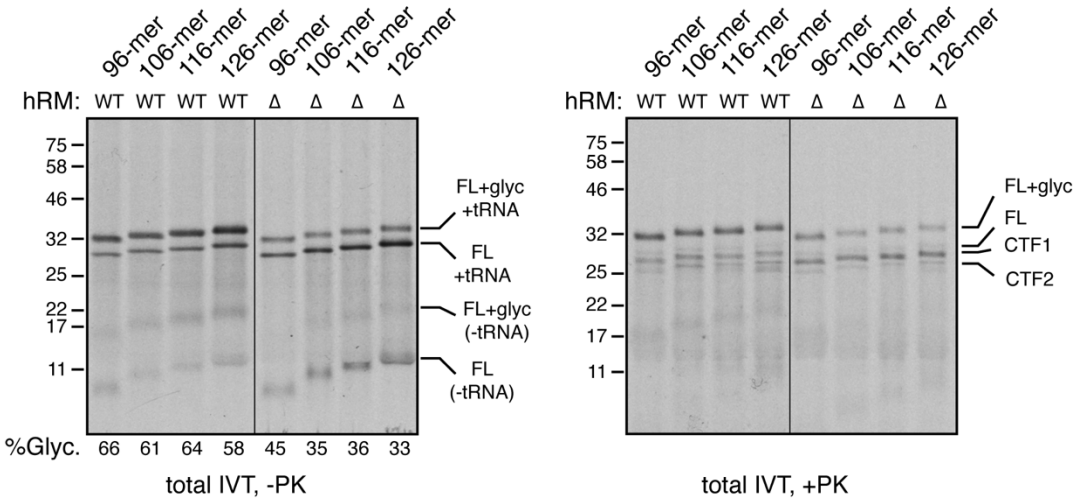


Figure 2.12 – EMC is required at early steps of β_1AR-TMD1 insertion. Ribosome-nascent chain complexes of gHA-β1AR truncated at the indicated lengths were produced in RRL, incubated with wild type (WT) or ΔEMC6 (Δ) hRM, and analyzed directly (total IVT, -PK) or subjected to digestion with proteinase K (PK) before analysis (total IVT, +PK). The products are labeled as in Fig. 2.11.

Chemical crosslinking of RNCs via a cysteine preceding the TMD validated the conclusions from the protease protection assay. In the crosslinking assay, we

monitored crosslinks between the nascent chain and a single cytosolic cysteine in Sec61β to assess the cytosolic disposition of sequences preceding the TMD (Fig. 2.13A). At each length, crosslinking to Sec61β was greater in ΔEMC6 microsomes than matched WT reactions (Fig. 2.13B, bottom panel). This is the mirror image of the extent of glycosylation in these same samples (top panel) because cysteine availability in the cytosol is mutually exclusive with glycosylation of an acceptor site four residues away. This indicates that RNCs that fail successful N_{exo} insertion in ΔEMC6 microsomes are at the Sec61 translocon with the N-terminus facing the cytosol. For the reasons articulated above, this difference is less prominent for the 96-residue RNCs. Taken together, the findings with β_1AR-TMD1 suggest that nascent β_1AR normally engages SRP, targets to the ER, and inserts in the N_{exo} orientation in a reaction that is stimulated by EMC. In the absence of EMC, N_{exo} insertion is less efficient resulting in the non-inserted β_1AR being observed near the Sec61 translocon, demonstrating topological inversion at the earliest point of TMD insertion.

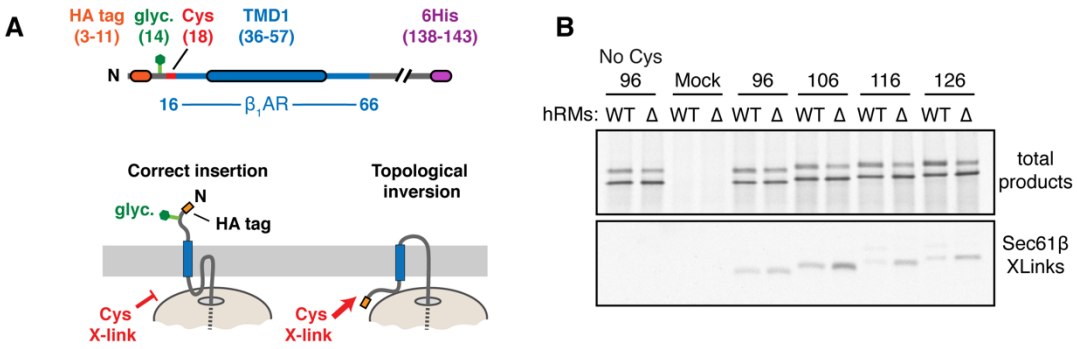


Figure 2.13 – Topological inversion occurs at early stages of β_1AR-TMD1 insertion. (A) Diagram showing position of N-terminal Cys located upstream of TMD1 and the interpretation of the chemical crosslinking assay. **(B)** ^{35}S-methionine labeled β1AR nascent chains of the indicated lengths were targeted to WT or ΔEMC6 hRMs (top panel), then subjected to sulfhydryl-mediated crosslinking. The crosslinked products were immunoprecipitated using antibodies against Sec61β and shown in the bottom panel. Controls lacking either mRNA (mock) or a cysteine in the nascent chain showed no Sec61β immunoprecipitated products. All immunoprecipitations were loaded such that the total represents 10% of the reaction and the IPs are 25% of the total reaction.

TMD1 of most GPCRs requires EMC for optimal insertion

To determine whether the first TMDs of other GPCRs also rely on EMC, we analyzed constructs containing TMD1 and flanking regions of sixteen GPCRs in a context similar to β_1AR-TMD1. Using glycosylation of an N-terminal site in 116-residue RNCs as the readout, we found that all GPCRs tested showed at least a partial dependence on EMC, ranging from ~20% to over 90% impairment in its absence (Fig. 2.14). Collectively, this sample of GPCRs represents a distribution of N_{exo} signal anchors with diverse sequences and biophysical properties such as hydrophobicity (TMD tendency) and flanking charge distribution (Table 2.1). However, there was no obvious unifying property between substrates that could easily provide an explanation for EMC dependence (Table 2.1 and Fig. 2.14). Consistent with the lack of effect in cells (Fig. 2.1A), ASGR1 showed little or no deficiency in insertion into ΔEMC6 microsomes, while LepB showed a very small but reproducible EMC-dependence (Fig. 2.14).

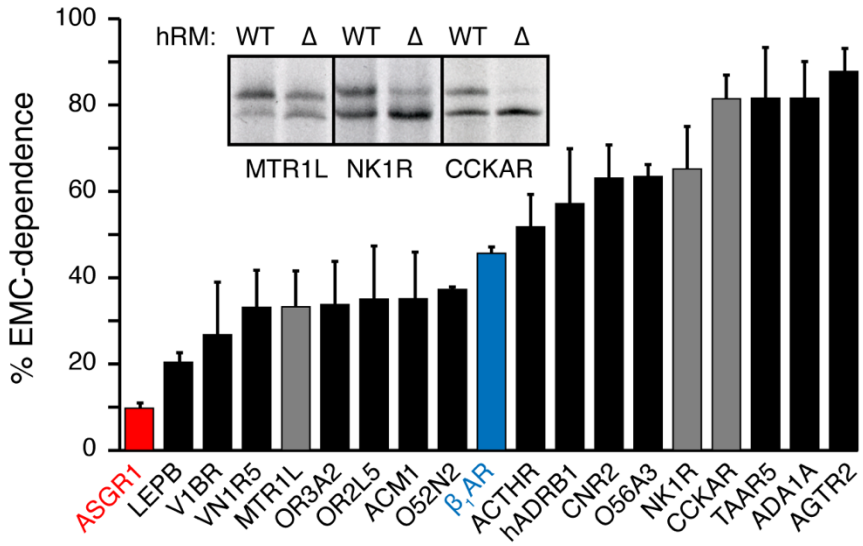


Figure 2.14 - TMD1 of most GPCRs requires EMC for optimal insertion. Constructs containing TMD1 and flanking regions from the indicated GPCRs (see Table S1) were analyzed by glycosylation of nascent chains targeted to wild type (WT) or ΔEMC6 (Δ) hRMs. The % decrease in ΔEMC6 hRM was quantified from three experiments and plotted and error bars represent standard deviation from the mean. Example data from the three GPCRs indicated by grey bars are shown in the inset. The model proteins ASGR1 and LepB were also analyzed for EMC-dependence and plotted for comparison. These experiments were aided by A. Guna.

name	sequence	hydro. (TM tend.)	charge Δ(C-N)
LepB	ANMFALILVIATLVTGILWSVDKFFFAPKRRERQAAAQ	21.9	+3
ASGR1	PPPPQPMLQRLSSGPRLLLLSLGLSLLLLVVVSVIGSQNSQLQEELR	23.5	-3
V1BR	ATTPWMGRDEELAKVEIGVLATVLVLATGGNLAVLLTLGQLGRKRSRM	17.1	+5
VN1R5	LKLVIIENMAEIMLFSLDLLLFSTDILSFNFPSKMIKLPG	13.7	0
MTR1L	GSIGSKLPQPEYPPALIIFMFSAMVITIVVDLIGNSMVILAVTKNKKL	17.7	+3
OR3A2	AEFILLGLVQTEEMQPVVFVLLLFAYLVTTGGNLSILAAVLVEPKLHA	21.2	+4
OR2L5	ILLGLFPPSKIGLFLFILFVLIFLMALIGNLSMILLIFLDTHLHTPMY	34.6	-2
ACM1	PQITVLAPGKGPWQVAFIGITTGLLSLATVTGNLLVLISFKVNTELKT	18.0	0
O52N2	GFFILNGVPGLEATHIWISLPFSFMYIIAVVGNSGLISLISHEEALHR	17.6	0
β_1AR	RQVSAELLSQQWEAGMSLLMALVVLLIVAGNVLVIAAIGSTQRLQTLT	21.1	+1
ACTHR	QNTARQNSDSPRVVMPEEIFFTISIVGVLENLIVLLAVFKNKNLQAPM	17.3	+1
ADRB1	PASESPEPLSQQWTAGMGLLMALIVLLIVAGNVLVIVAIAKTPRLQTL	28.0	0
CNR2	SNPMKDYMILSGPQKTAVAVLCTLLGLLSALENVAVLYLILSSHQLRR	20.0	+1
O56A3	NSFVRSPSWQHWLSLPLSLLFLLAVGANTTLLMTIWLEASLHQPLYYL	17.2	-1
NK1R	TSEPNQFVQPAWQMVLWAAAYTVIVVTSVVGNVVVMWIILAHKRMRTV	20.5	+4
CCKAR	SLDQPRPSKEWQPAMQILLYSLIFLLSVLGNTLVITVLIRNKRMRTVT	25.8	+4
TAAR5	VQGSSPRTVHTLGIQLVIYLASAAGMLIIVLGNVFVAFAVSYFKALHT	21.7	0
ADA1A	SQSTQPPAPVQMSKAILLGVILGGLILFGVLGNILVILSVASHRHLHS	27.2	0
AGTR2	STLQSSQKPSDKHMDAIPILYYIIFVIGFLVNIVVVTLFSSQKGPKKV	27.3	+3

Table 2.1 - Sequences of TMD regions analyzed in this study (related to Fig. 2.14). The GPCR-TMD1 cassette depicted in Fig. 3A was populated with the first TMD and adjacent flanking regions (~10 residues on either side) of the indicated GPCRs. The TMD regions of the model proteins bacterial leader peptidase (LepB) and ASGR1 are shown for comparison. The TMD region (as annotated by Uniprot) is underlined, and flanking basic and acidic residues are shown in red and blue, respectively. The calcuated TM tendency score and charge difference are indicated for each TMD region. Note that the Uniprot annotated TMD may not be precise in defining the boundaries, so parameters such as TMD length and hydrophobicity should be interpreted with this caveat in mind.

Conclusions from glycosylation analyses were verified by protease protection assays and N-terminal immunoprecipitation (three examples are shown in (Fig. 2.15). Importantly, analysis of RNC intermediates for three native GPCR N-terminal sequences showed a similar degree of impaired insertion in ΔEMC6 microsomes as seen for the respective epitope-tagged TMD1 constructs (Fig. 2.15). This observation not only validates earlier assays in the assessment of EMC dependence using tagged TMD1 variants, but also suggests the sequence information imposing EMC dependence is encoded within our simplified constructs, i.e. the TMD and 10 flanking N- and C-terminal residues (Table 2.1). Thus, early events in the biogenesis of most GPCRs differs at least partially in EMC-deficient microsomes.

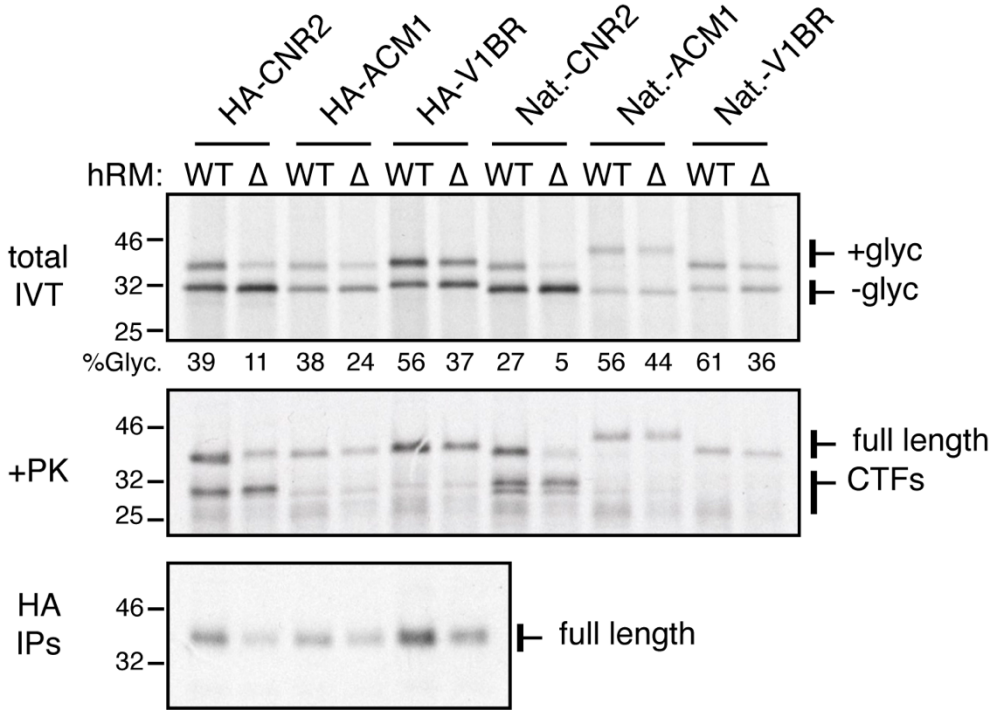


Figure 2.15 – EMC is required for early events in GPCR biogenesis. Ribosome-nascent chains (stalled ~60 residues downstream of the indicated TMDs) were targeted to wild type (WT) or ΔEMC6 (Δ) hRMs and analyzed by the PK-protection assay as in. 'HA' indicates an N-terminal HA tag and glycosylation site (see Fig. 2.9A), while 'Nat.' indicates the native N-terminal domain. The PK-digested samples from the HA-containing constructs were also subjected to immunoprecipitation (HA IPs). All immunoprecipitations were loaded such that the total represents 10% of the reaction and the IPs are 25% of the total reaction.

Three full-length GPCRs were analyzed in U2OS cells for impaired biogenesis using the dual-color flow cytometry assay (Fig. 2.16). As expected for expression in a heterologous cell type, the steady state levels of each receptor varied somewhat, with the Type 2 angiotensin II receptor (AGTR2) expressing better than either α_{1A}-adrenergic receptor (ADA1A) or Cannabinoid receptor 2 (CNR2). Nevertheless, the steady state level of the GFP-tagged GPCR was reduced in ΔEMC5 cells for each protein, but restored to WT levels when EMC5 was re-expressed. Thus, the insertion impairment of TMD1 seen in ΔEMC6 microsomes in vitro corresponds to reduced post-translational stability of the full GPCR in ΔEMC5 cells.

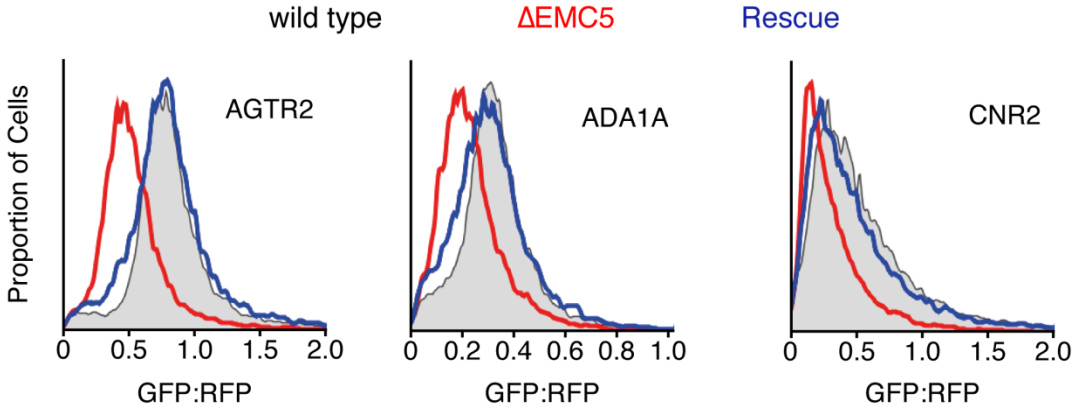


Figure 2.16 – EMC is required for full-length GPCR biogenesis in cells. The indicated GPCRs were tagged as in Fig. 2.1A and analyzed by flow cytometry as in Fig. 2.1. Gray trace is WT cells, red trace is ΔEMC5 cells, and blue trace is EMC5-rescued ΔEMC5 cells. The above experiment was performed with help from S. Juszkiewicz.

Prior studies analyzing sequence properties such as charge bias, hydrophobicity, and TMD length demonstrated relatively modest effects, of each individual feature, on topology. More puzzling was the machinery responsible for actually decoding these properties in order to bias orientation. Indeed, extensive mutagenesis of Sec61 revealed only modest effects on topology, and only on already topologically compromised substrates (Goder et al., 2004; Junne et al., 2007). Nevertheless, general trends demonstrate C-terminally flanking positive charges, increasing hydrophobicity, and short TMDs favor N_{exo} insertion of a signal anchor (Higy et al., 2004). Similar analyses of β_1AR-TMD1 mutants show a link between TMD topology and EMC dependence (Fig. 2.17A, Table 2.2), with reduced length, increased

hydrophobicity, and flanking charges that favor a cytosolic C-terminus all reducing EMC dependence. Conversely, LepB could be made highly EMC dependent by lengthening its TMD with three non-hydrophobic residues (Fig. 2.17B-C, Table 2.2). These observations partially explain the variable EMC dependence observed amongst the natural signal anchors of GPCRs and also correlatively link EMC to topological determination. However, how these biophysical features are recognized and what determines the fate of a signal anchor will require more in-depth analysis.

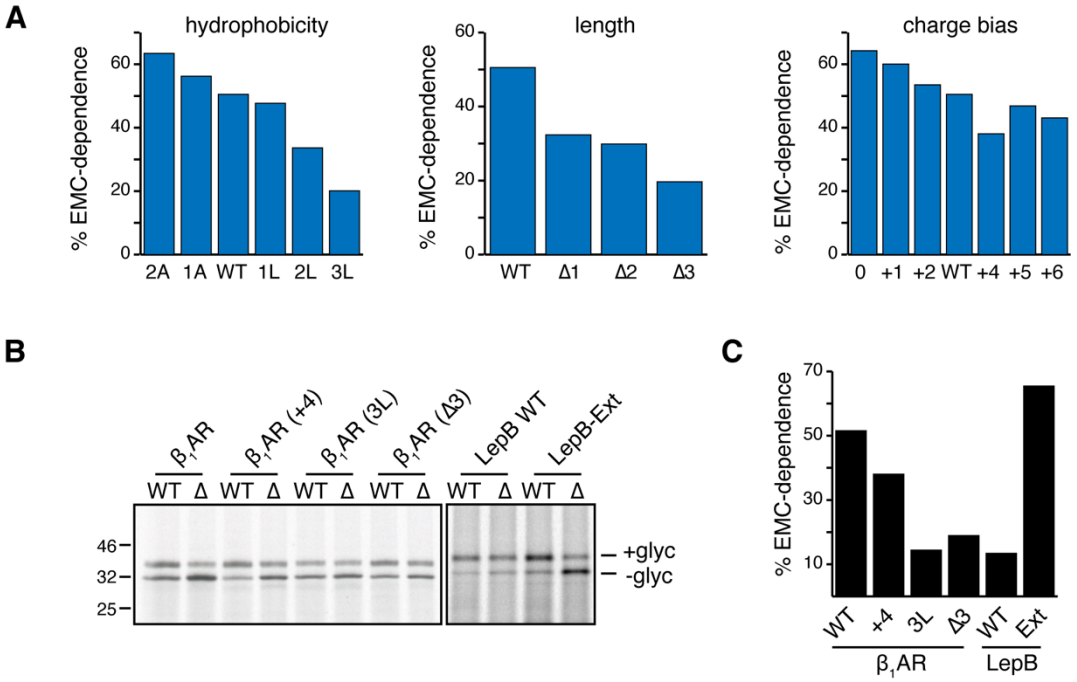


Figure 2.17 - Properties of TMD influencing EMC dependence. (A) The indicated β1AR constructs (see Table 2.2) were tested for insertion into wild type and ΔEMC hRM as in Figure 2.15-2.16. Glycosylation was used to quantify the amount of insertion in the correct (Nexo) orientation. The relative difference in correct insertion between wild type and ΔEMC microsomes was used to determine EMC-dependence (i.e., 60% insertion in ΔEMC relative to wild type would mean 40% EMC-dependence). All of the constructs were analyzed together. The wild type is re-plotted in each of the three graphs for comparison. Note that EMC-dependence of the β1AR TMD is influenced by hydrophobicity, TMD length, and to a lesser extent, flanking charge bias. **(B, C)** Ribosome-nascent chains of the indicated constructs (Table 2.2) were analyzed for insertion by the glycosylation assay using WT and ΔEMC6 hRMs. Panel C shows quantification of the autoradiograph shown in panel B.

name	sequence	hydro. (TM tend.)	charge Δ(C-N)
β_1AR	QVSAELLSQQWEAGMSLLMALVVLLIVAGNVLVIAAIGSTQRLQTLT	26.3	+3
1A	QVSAELLSQQWEAGMSLLMAAVVLLIVAGNVLVIAAIGSTQRLQTLT	24.9	+3
2A	QVSAELLSQQWEAGMSLLMAAVVALIVAGNVLVIAAIGSTQRLQTLT	23.4	+3
1L	QVSAELLSQQWEAGMSLLMALVVLLIVLGNVLVIAAIGSTQRLQTLT	27.7	+3
2L	QVSAELLSQQWEAGMSLLMALVVLLIVLLNVLVIAAIGSTQRLQTLT	29.8	+3
3L	QVSAELLSQQWEAGMSLLMALVVLLIVLLLVLVIAAIGSTQRLQTLT	33.2	+3
0	QVSAQLLSQQWQAGMSLLMALVVLLIVAGNVLVIAAIGSTQQLQTLT	26.3	0
+1	QVSAQLLSQQWQAGMSLLMALVVLLIVAGNVLVIAAIGSTQRLQTLT	26.3	+1
+2	QVSAELLSQQWQAGMSLLMALVVLLIVAGNVLVIAAIGSTQRLQTLT	26.3	+2
+4	QVSAELLSQQWEAGMSLLMALVVLLIVAGNVLVIAAIGRTQRLQTLT	26.3	+4
+5	QVSAELLSQQWEAGMSLLMALVVLLIVAGNVLVIAAIGRTQRLKTLT	26.3	+5
+6	QVSAELLSEQWEAGMSLLMALVVLLIVAGNVLVIAAIGRTQRLKTLT	26.3	+6
Δ1	QVSAELLSQQWEAGMSLLMALVVLLIVAG-VLVIAAIGSTQRLQTLT	27.9	+3
Δ2	QVSAELLSQQWEAGMSLLMALVVLLIVA--VLVIAAIGSTQRLQTLT	28.1	+3
Δ3	QVSAELLSQQWEAGMSLLMALVVLLIV---VLVIAAIGSTQRLQTLT	27.7	+3
LepB	ANMFALILVIATLVTGILWSVDKFFFAPKRRERQAAAQ	21.9	+3
LepB-Ext	ANMFALSILVGIATLVTGILTWSVDKFFFAPKRRERQAAAQ	20.8	+3

Table 2.2 - Sequences of TMD mutants analyzed in this study (related to Fig. 2.16). The β1AR-TMD1 and LepB constructs were mutated as indicated (green residues indicate changes). The calcuated TM tendency score and charge difference are indicated for each TMD region. The TMD is underlined. Note that the assignment of the TMD for β1AR is different from that indicated in Uniprot (Table S1) and is based on the known structure of β1AR. Although not shown here, we have verified that the effect of 3L and Δ3 are due to the increase in hydrophobicity and decrease in TMD length, respectively, and not to the specific residues that are mutated. This was done by mutating or deleting three other residues in the TMD to achieve the same approximate hydrophobicity and length.

N_{exo} signal anchor insertion can occur without the Sec61 complex

Despite the clear relationship between established topological determinants and EMC dependence (Fig 2.17), analysis of native GPCR TMDs argued that competing biophysical properties could partially bypass EMC necessity (Fig. 2.14, Table 2.1). The TMDs that display EMC-dependence (whether partial or near-complete) are diverse but the unifying feature of these proteins is their N_{exo} topology. By contrast, proteins whose targeting elements acquire the N_{cyt} topology at the membrane are EMC-independent: the cleavable signal peptides of prolactin and PrP, and the signal anchor of ASGR1. This correlation suggested the hypothesis that EMC might use its insertase activity (Guna et al., 2017) to co-translationally insert N_{exo} signal anchors, while the Sec61 complex accommodates N_{cyt} targeting elements (Voorhees and Hegde, 2016). In support of this idea, it is noteworthy that tail-anchors inserted by EMC are similar to N_{exo} signal anchors in having relatively short translocated domains (Fig. 2.18A) and basic residues enriched on the cytosolic flank of the TMD (Fig. 2.18B).

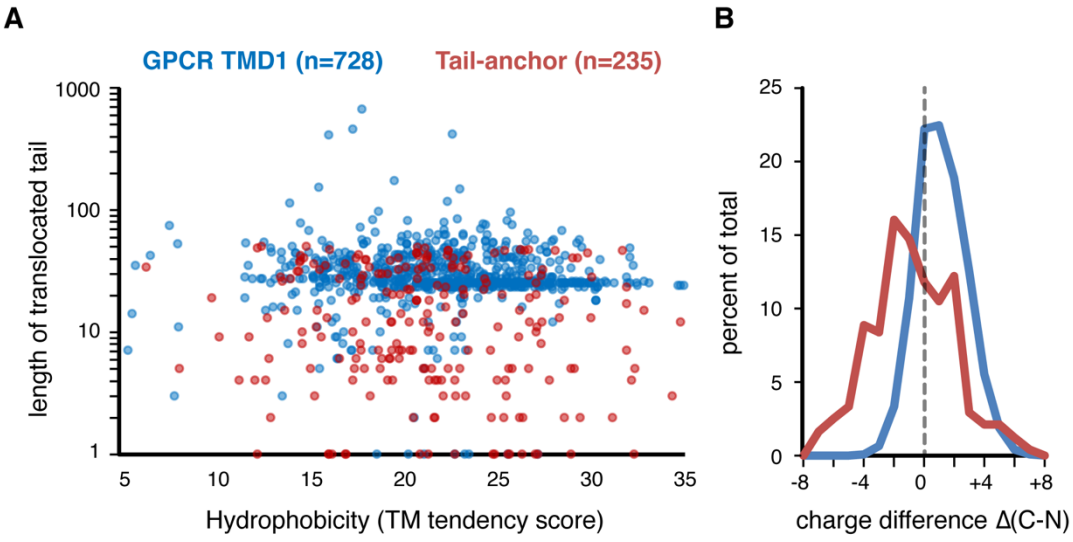


Figure 2.18 – The shared properties of EMC dependent substrates. **(A)** Plot of TM tendency score versus length of the translocated domain for all non-signal-containing GPCRs and ER-localized TA proteins in the human genome. The translocated domain of almost all TA and N_{exo} signal anchors is less than ~40 residues. **(B)** Histogram of the charge difference for the dataset in panel A. Charge difference is calculated based on the sum of the C-terminal charges subtracted by the sum of N-terminal charges Δ(C-N). Note that in both cases, there is a slight preference for net positive charges facing the cytosol.

To investigate this hypothesis, we examined GPCR TMD1 insertion into membranes depleted of the Sec61 complex. RMs were solubilized, mock- or immunodepleted of the Sec61 complex, and the remaining proteins reconstituted into proteoliposomes (PLs) by slow removal of detergent. Sec61 was verified to be thoroughly depleted (by over 95%; Fig. 2.19A), while the overall protein profile was otherwise unchanged (Fig. 2.19B). As shown previously (Görlich and Rapoport, 1993), Sec61-depleted PLs are completely deficient in prolactin translocation (Fig. 2.19C). Furthermore, they cannot detectably insert the N_{cyt} signal anchored protein ASGR1 as measured by protease protection assays. Remarkably however, the N_{exo} signal anchor from β_1AR can be inserted into to Sec-depleted proteopliposomes with near equal efficiency (Fig. 2.19C).

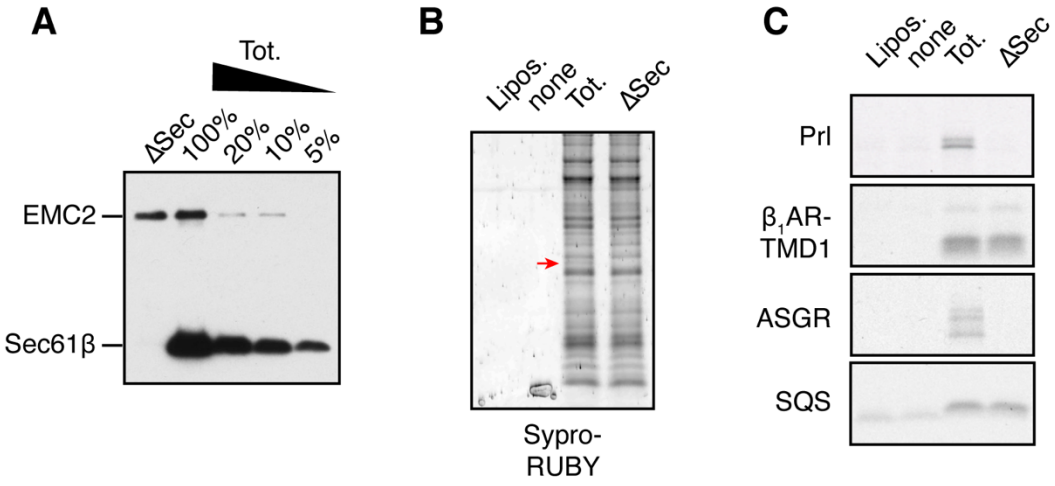


Figure 2.19 – β_1AR N_{exo} signal anchor insertion is Sec61-independent. (A) Immunoblotting of proteoliposomes (PLs) reconstituted from total ER proteins (Tot.) or Sec61-depleted ER proteins (ΔSec) shows that under conditions where even 5% of total PLs show readily detectable Sec61, none is seen in ΔSec PLs. EMC levels are comparable. Verifying that Sec61 was depleted with >95% efficiency **(B)** Total proteins in the proteoliposomes were visualized by Sypro Ruby, with the position of Sec61a indicated by the red arrow. **(C)** The protease-protected (and hence, translocated) products were recovered by immunoprecipitation using the respective tags Prl (Ab against native Prl N-term), β_1AR-TMD1 (HA tag), ASGR (HA tag), and SQS (3F4 tag) to verify insertion/translocation. These experiments were performed with help from R.S. Hegde.

Additionally, several different N_{exo} signal anchored proteins from GPCRs can be inserted into Sec61-depleted PLs. Protease-protection assays and IPs via an N-terminal tag showed that the N-terminus is protected from digestion in mock- and Sec61-depleted PLs, but not empty liposomes (Fig. 2.20). Importantly, the extensively studied model protein LepB whose insertion was thought to require the Sec61 complex (Heinrich et al., 2000) was inserted almost equally well in non-depleted or depleted PLs (Fig. 2.20).

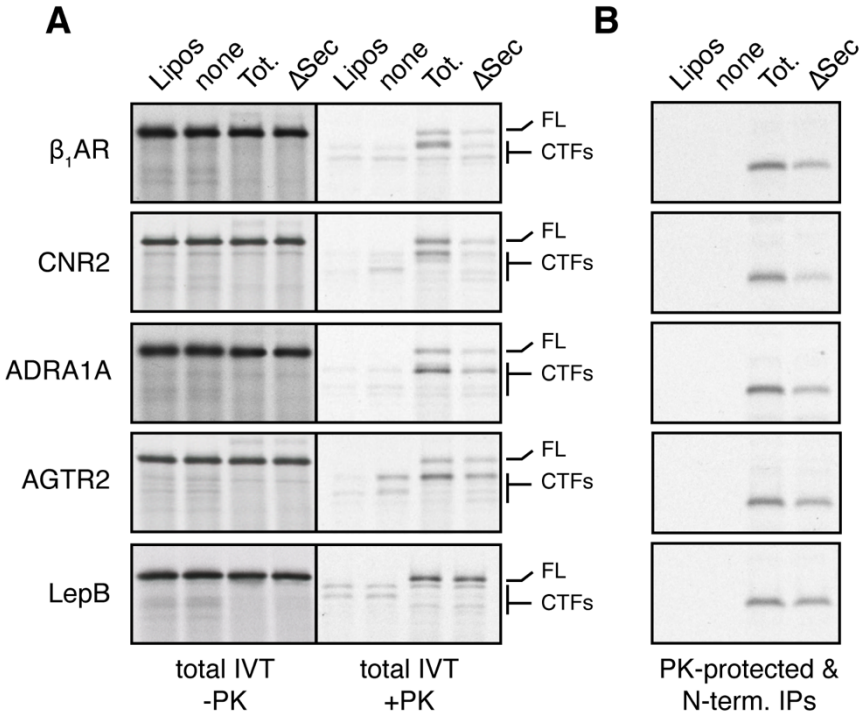


Figure 2.20 – Multiple N_{exo} signal anchors insert independently of Sec61. (A) Ribosome-nascent chain complexes of constructs containing the indicated TMD1 regions (see Table 2.1, 2.2) truncated ~60 residues beyond the TMD (corresponding to residue 116 in the β1AR-TMD1 construct) were produced in RRL. They were incubated without anything, with liposomes, or with PLs from total ER proteins (Tot.) or Sec61-depleted ER proteins (ΔSec). An aliquot of the sample was analyzed directly (-PK) or subjected to digestion with proteinase K (+PK). **(B)** An aliquot of the PK-digested sample was subsequently immunoprecipitated via the N-terminal HA tag after RNAse digestion (N-term. IPs). FL indicates full length product protected from protease, indicative of successful insertion. CTFs indicate C-terminal fragments from non-inserted products. All immunoprecipitations were loaded such that the total represents 10% of the reaction and the IPs are 25% of the total reaction. These experiments were performed with help from R.S. Hegde.

Although glycosylation is relatively inefficient in PLs, over-exposed autoradiographs showed that an N-terminal acceptor site is glycosylated comparably efficiently for several different N_{exo} signal anchors in both mock- and Sec61-depleted PLs (Fig. 2.21). This observation is important for two reasons. First, it provides an orthogonal verification of correct insertion and the fidelity of downstream enzymatic events acting on the nascent chain, effectively further validating the reconstituted system. Second, glycosylation establishes productive translocation of the N-terminus in to the ER lumen and confirms the interpretations from assays using protease protection and N-terminal IPs (Fig. 2.19-20).

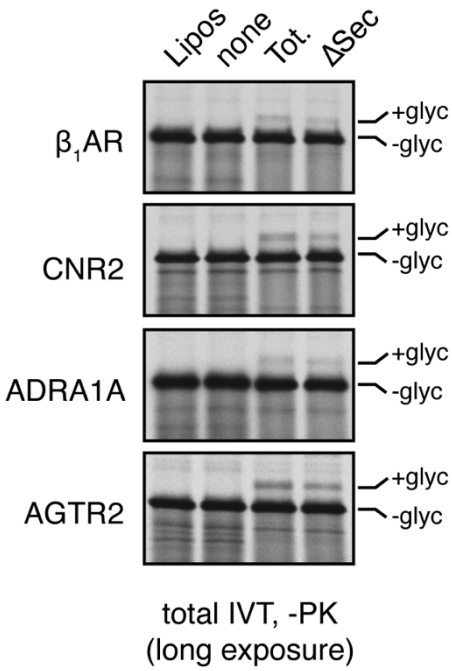


Figure 2.21 – Accurate insertion of N_{exo} signal anchor insertion independently of Sec61. The total IVT products from Fig. 2.20 shown from an overexposed autoradiograph to visualize the minor glycosylated product (+glyc). Glycosylation is relatively inefficient in PLs compared to native microsomes. These experiments were performed with help from R.S. Hegde.

While insertion of some of these signal anchors was reduced by ~50% upon Sec61 depletion, others were essentially unaffected. As discussed below, this reduction may be due to an inability of ribosomes to stably dock at the membrane in the absence of Sec61 (Kalies et al., 1994). Despite this limitation, the data illustrate that N_{exo} signal anchors do not strictly require Sec61 for insertion, in stark contrast to a signal peptide or N_{cyt} signal anchor. In light of this result, it is noteworthy that N_{exo}

signal anchors are the only class of substrates completely resistant to a potent Sec61 inhibitor (McKenna et al., 2017) that prevents opening of the Sec61 channel by signals and TMDs (Baron et al., 2016; MacKinnon et al., 2014). Both of these observations can be explained by a model where EMC, not Sec61, plays a primary role during insertion of N_{exo} signal anchors.

EMC is sufficient for N_{exo} signal anchor insertion

To test whether EMC's insertase function can explain Sec61-independent insertion of N_{exo} signal anchors, we prepared PLs containing purified EMC without or with SRP receptor and tested their capacity for translocation and membrane insertion. Inclusion of SR in the proteopliposomes improves targeting of RNCs and greatly enhances the efficiency of insertion in a purified system (Görlich and Rapoport, 1993). Importantly, we verified that EMC and EMC/SR PLs are not contaminated with any detectable Sec61 complex (Fig. 2.22A-B). Consistent with a strict requirement for Sec61 complex, neither prolactin nor ASGR1 showed detectable translocation in EMC or EMC/SR PLs. By contrast, the β_1AR-TMD1 was inserted into EMC-containing PLs. The additional presence of SR stimulated insertion of β_1AR-TMD1, but not of the tail-anchored protein squalene synthase (Fig. 2.22C). It is likely that this stimulation is due to SR facilitating dissociation of the TMD from SRP. SR alone had no translocation or insertion activity, suggesting that simply delivering β_1AR-TMD1 to the membrane surface is insufficient to allow insertion. Importantly, EMC mediated insertion in the purified system was observed with either a terminated nascent chain, or an RNC in which β_1AR-TMD1 is still tethered to the ribosome (Fig. 2.22D). The latter represents a situation in which unregulated insertion would be highly unlikely due to the biophysical constraints of being tethered to the extremely large ribosome. The overall lower insertion efficiencies into these purified EMC PLs relative to PLs containing total ER proteins is probably explained by the lower level of EMC in the purified system (Fig. 2.22A-B) and the absence of a ribosome docking site normally provided by the Sec61 complex (Kalies et al., 1994). These limitations notwithstanding, we conclude that after targeting, EMC is sufficient to mediate insertion of not only tail-anchored proteins like SQS, but also N_{exo} signal anchors.

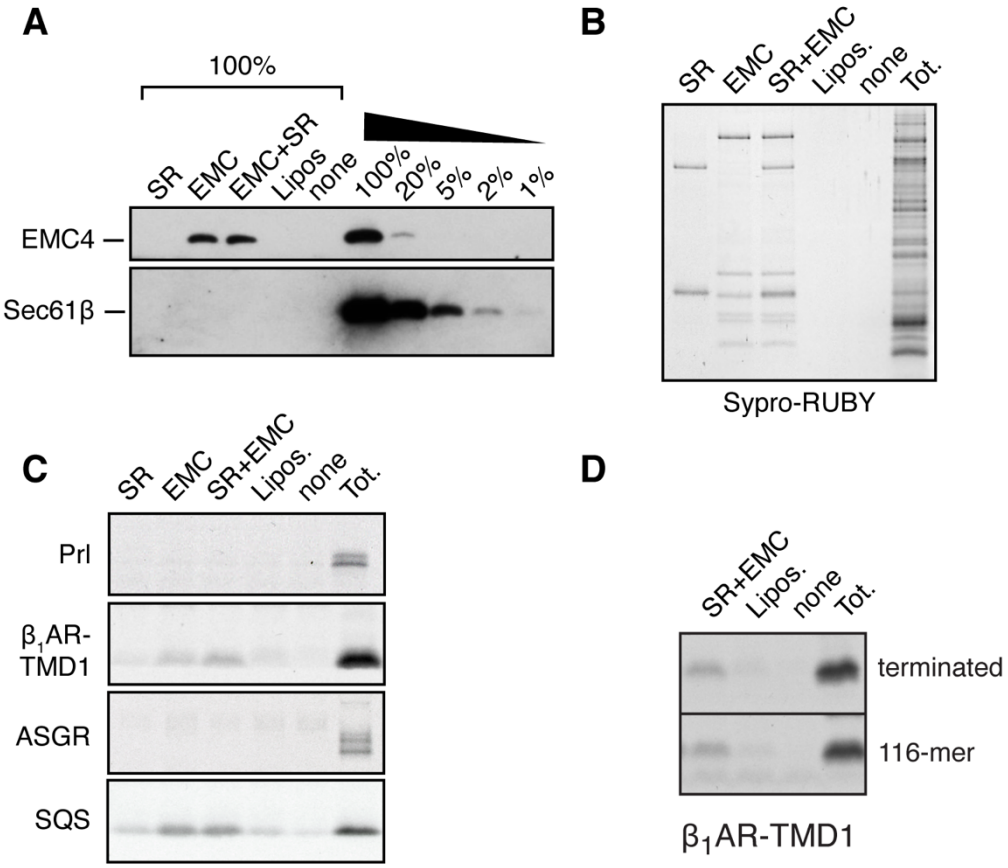


Figure 2.22 – EMC is sufficient for N_{exo} signal anchor insertion. (A) The indicated PLs were analyzed by immunoblotting for Sec61 and EMC to verify no Sec61 contamination of either EMC or SRP receptor (SR) PLs. The serial dilution on the right-hand side of the gel represents total proteoliposomes reconstituted in parallel **(B)** Proteins in the proteoliposomes were visualized by Sypro Ruby. 10-fold excess of the first four lanes were loaded to detect the purified proteins. **(C)** The indicated constructs were analyzed by the protease-protection assay for translocation and immunoprecipitated. Abbreviations: liposomes (Lipos), proteoliposomes reconstituted from total ER proteins (Tot.), and proteoliposomes containing the indicated purified proteins (SR is SRP receptor). **(D)** Insertion assay as in (C) into the indicated proteoliposome preparation. The terminated β1AR-TMD1con-struct (as in C) was compared to the same construct stalled at residue 116 (~60 residues downstream of the TMD). Shown is the immunoprecipitated protease-protected N-terminal fragment diagnostic of successful in-sertion in the Nexo topology. Note that specificity and efficiency of insertion is comparable for the terminated and stalled versions of β1AR-TMD1. These experiments were performed with help from R.S. Hegde.

EMC and Sec61 can function sequentially to insert two TMDs

The findings thus far indicate that EMC is needed for efficient insertion of TMD1 of GPCRs in the N_{exo} topology, that Sec61 complex is not strictly required for this

step, and that purified EMC is sufficient in a reconstituted system to mediate N_{exo} signal anchor insertion. In the context of a full length GPCR, the next step after TMD1 insertion is TMD2 insertion in the opposite orientation. The reconstitution experiments with ASGR1 indicate that co-translational insertion in this topology requires Sec61 and cannot be mediated by EMC.

To test whether TMD2 of β_1AR requires Sec61, we analyzed a two-TMD construct (Fig. 2.23 and 2.24) for insertion in reconstituted PLs containing or lacking the Sec61 complex. We first characterized insertion of TM1+2 in native RMs using a construct containing glycosylation sites and epitope tags at both the N- and C-terminus (Fig. 2.23A-B). These initial analyses showed that its insertion in the correct double-spanning topology results in the addition of two glycans and a protein that is fully shielded from cytosolic protease due to the inaccessibility of the short intervening cytosolic loop (Fig. 2.23C, and diagram in Fig. 2.23B). The identity of each molecular weight species was confirmed by inhibition of glycosylation with saturating amounts of acceptor peptide (AP) and immunoprecipitations via either terminus. Polypeptides that fail insertion entirely are digested by cytosolic protease, while those with only the first TMD inserted in the N_{exo} topology generate a protected N-terminal fragment that can be recovered by an N- but not C-terminal IP (Fig. 2.23C).

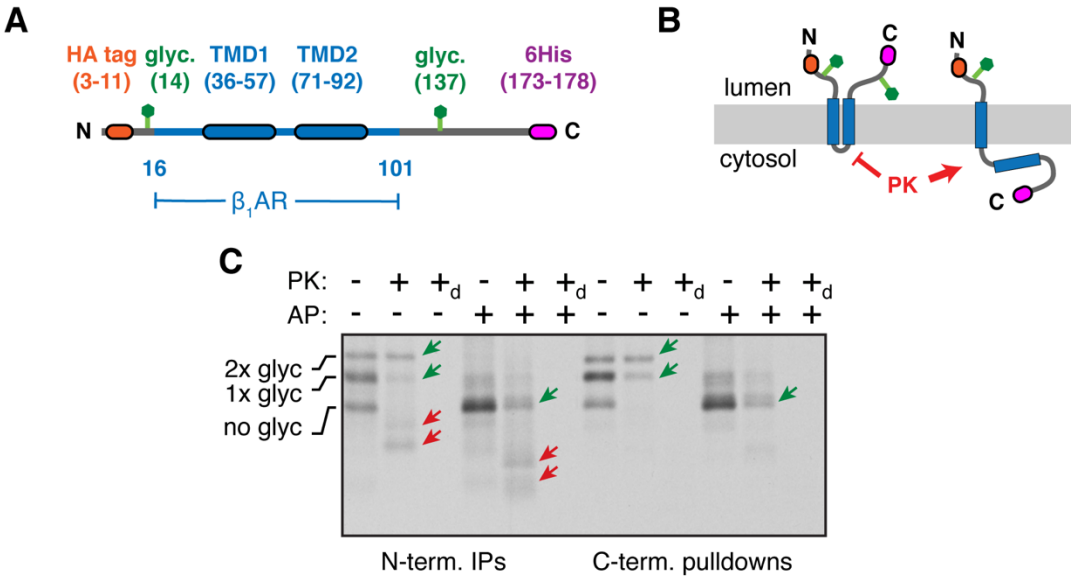


Figure 2.23 – Characterization of TMD1+2 insertion in ER rough microsomes. (A) Diagram of the two-TMD β1AR construct (β1AR-TMD1-2), **(B)** and its topology when TMD2 inserts or fails to insert into the membrane. Only the single-spanning form would be accessible to proteinase K (PK) digestion due to the short loop between TMD1 and TMD2. In addition, the double-spanning topology can be glycosylated twice, while the single spanning topology is only glycosylated once. (C) 35S-methionine labeled β1AR-TMD1-2 was translated in reticulocyte lysate (RRL) in the presence of canine pancreas-derived rough microsomes. Where indicated, the translation reaction contained an acceptor peptide (AP) inhibitor of N-linked glycosylation. The translation products were either left untreated or digested with proteinase K without or with detergent (subscripted d) as indicated. The samples were divided in two and recovered via the N- or C-terminal tag and analyzed by SDS-PAGE and autoradiography. The positions of unglycosylated, singly-glycosylated (1x glyc) or doubly-glycosylated (2x glyc) products are indicated. Green arrows indicate products that are fully protected from protease digestion and represent the double-spanning topology. Red arrows indicate N-terminal protease-protected fragments. Some heterogeneity is observed in the size of these fragments presumably due to heterogeneity in where the protease digestes the exposed polypeptide.

To determine whether Sec61 was necessary for insertion of the second TMD in β_1AR we turned to a simplified construct, in which the second C-terminal glycosylation site had been removed. This variant behaved identically by protease protection assays but provided a simplified readout for direct comparison in proteoliposomses (Fig. 2.24A), which have substantially reduced glycosylation efficiencies. Little or no protease-protection was observed in reactions containing empty liposomes, or if the PLs were added post-translationally to the reaction (Fig. 2.24B). Notably, insertion of the first TMD occurred in the absence of Sec61, generating the N-terminal protected fragment that could be immunoprecipitated solely by the N-terminal epitope tag (Fig 2.24B and Fig. 2.24C). Insertion in the correct double-spanning topology was observed in mock-depleted PLs, but sharply reduced in Sec61-depleted PLs (Fig. 2.24B and Fig. 2.24C). Thus, co-translational topogenesis of the first two TMDs of β_1AR requires Sec61. The point at which Sec61's role becomes critical is TMD2 insertion, as TMD1 insertion can proceed in its absence. TMD1 insertion can be mediated solely by EMC (Fig. 2.22), although EMC's absence is partially tolerated by β_1AR presumably because its insertion by Sec61 occurs in the correct orientation for a subset of molecules. This indicates that although the correct double-spanning topology can be achieved without EMC, optimal topogenesis requires the combined functions of EMC and Sec61 for insertion of TMD1 and TMD2, respectively.

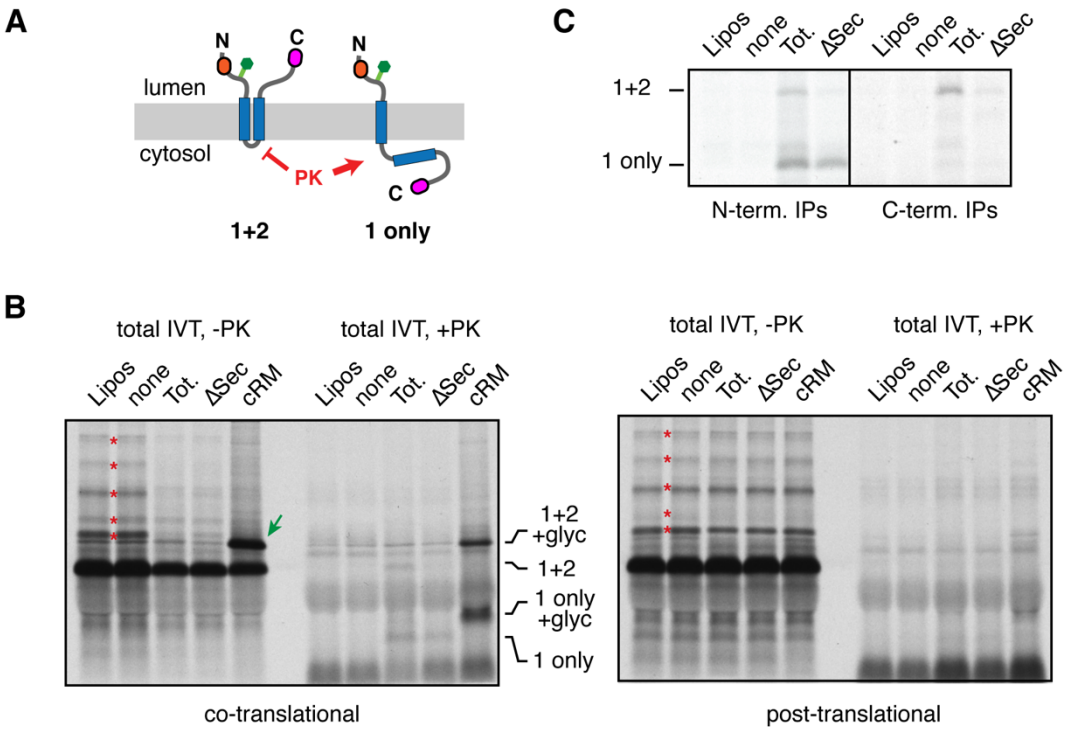


Figure 2.24 – EMC and Sec61 cooperate to insert multiple TMDs. **(A)** Diagram of the two-TMD β1AR construct and its topology when TMD2 inserts or fails to insert into the membrane. Only the single-spanning form is accessible to proteinase K (PK) digestion. The two-TMD construct from panel C was analyzed in the indicated proteoliposome preparations by the protease-protection assay. **(B)** The two-TMD β1AR construct (as shown in A) was analyzed in the indicated proteoliposome preparations or canine-pancreas derived microsomes (cRM) by the protease-protection assay. Samples were analyzed directly without immunoprecipitation. The left panel shows the experiment when membranes are present during the translation reaction (co-translational), while the right panel shows the experiment when incubation with membranes was post-translational. Red asterisks indicate ubiquitinated products, green arrow indicates the glycosylated product, "1+2" indicates the protected product indicative of the double-spanning topology, and "1 only" indicates the single-spanning topology. **(C)** The panel shows the PK-digested products shown in (B) after recovery via N- or C-terminal tags as indicated. "1+2" indicates the protected product indicative of the double-spanning topology, and "1 only" indicates the single-spanning topology.

Bypass of EMC dependence by constraining TMD1 topology

The biochemical analyses using simplified N-terminal regions of β_1AR show that one explanation for the observed requirement for EMC in cells (Fig. 2.1) is its role in topogenesis of TMD1. To investigate whether EMC is required for insertion, folding, or maturation steps beyond TMD1 insertion, we designed versions of β_1AR whose TMD1 would necessarily insert via Sec61. Sec61 is both necessary and sufficient for signal sequences and N_{cyt} signal anchors to initiate translocation

without any appreciable role for EMC. We therefore extended the N-terminus of β_1AR with either a cleavable signal sequence and the secreted protein lysozyme (termed SS-T4L- β_1AR; see diagram, Fig. 2.25A) or a signal anchor from mannosidase I with a short linker (termed ManI- β_1AR). Both of these extensions should mediate targeting, initiation of translocation, and commitment of protein topology before TMD1 emerges from the ribosome. Because the polypeptide at this stage would be threaded within the Sec61 channel, TMD1 will enter Sec61 and can insert via its lateral gate in the correct orientation, thereby bypassing EMC's insertase function.

In vitro translocation and protease protection analysis of SS-T4L- β_1AR and ManI- β_1AR showed that its insertion occurs similarly in WT and ΔEMC6 RMs under conditions where β_1AR insertion is significantly impaired (Fig. 2.25A). Analysis in cells using the dual-color fluorescent reporter assay showed no difference in either SS-T4L- β_1AR or ManI- β_1AR between WT and ΔEMC5 cells (Fig. 2.25B). SS-T4L similarly rescued the EMC-dependence of AGTR2 and ADA1A (Fig. 2.25B).

This result has three important implications. First, it strongly argues against any indirect effects of EMC on GPCR levels. Hence, explanations such as globally altered trafficking, degradation, or other general perturbations leading to the reduced GPCR levels (as seen in Figs. 2.1 and 2.14) seem highly unlikely. Second, the biochemically demonstrated EMC-dependent step of TMD1 insertion characterized in vitro must be the mechanistic explanation for reduced GPCR levels in ΔEMC cells observed in vivo. Third, the insertase function of EMC used for TMD1 topogenesis appears to be the only step during GPCR biogenesis where EMC is required. Thus, we conclude that EMC's role in facilitating the biogenesis of many GPCRs is due to its requirement during TMD1 insertion in the N_{exo} topology.

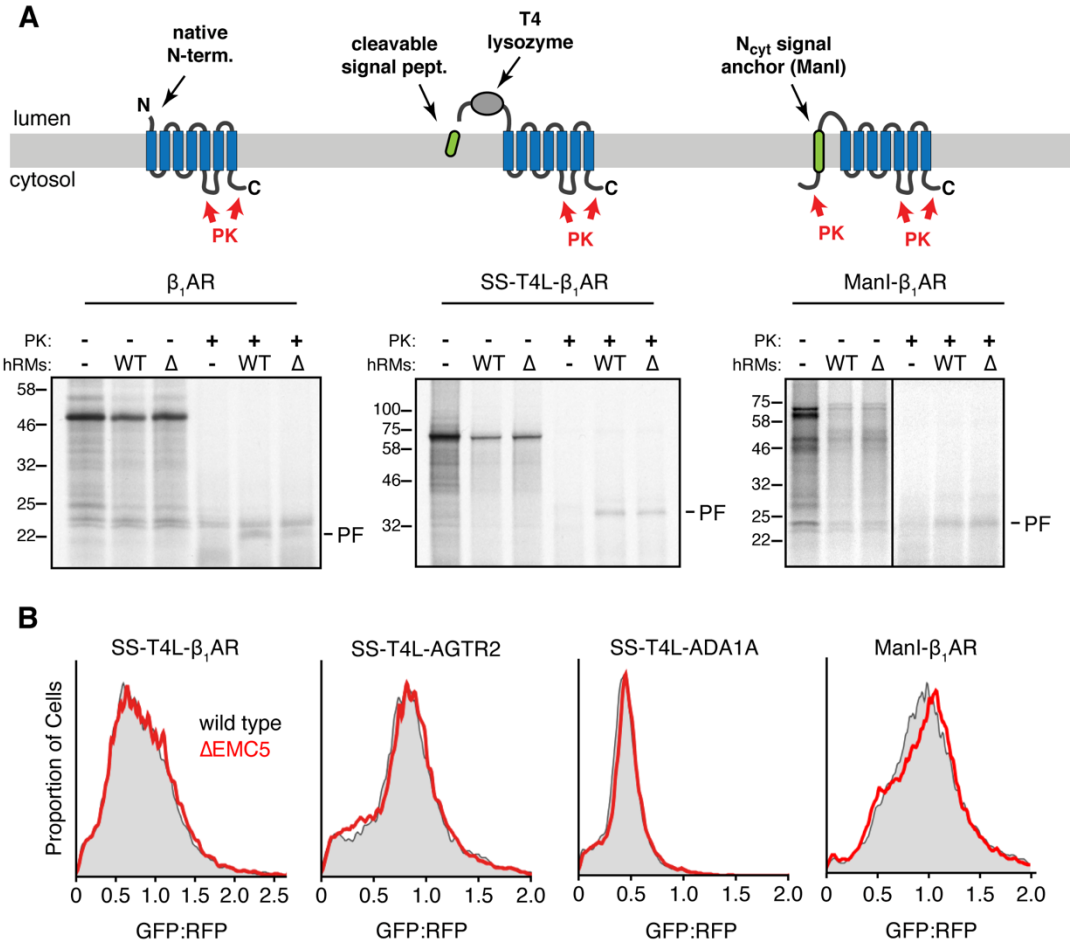


Fig. 2.25 - A Sec61-targeted signal sequence or TMD can bypass EMC-dependence in vitro and in vivo. (A) Diagram comparing the β1AR, SS-T4L-β1AR, and ManI-β1AR constructs (top) and their analysis of insertion into wild type (WT) or ΔEMC6 (Δ) hRM as in Fig. 2.4. PF indicates the protected N-terminal fragment generated by digestion of successfully inserted protein at the loop between TMD5 and TMD6 (see diagrams). (B) Flow cytometry analysis of the indicated constructs in wild type or ΔEMC5 U2OS cells as in Fig. 2.1. Note that in contrast to the matched constructs lacking the SS-T4L or ManI domains (Figs. 2.1 and 2.16), no appreciable consequence of EMC deletion is observed. The experiments in panel B were performed in parallel with the experiments shown in figure 2.1B as a single experiment, although they are displayed in separate figures. Please refer to figure 2.1B for the positive control in which β1AR-GFP-2A-RFP shows EMC dependence. The experiments in panel B were performed with help from S. Juszkiewicz.

A model for EMC mediated co-translational insertion

We have identified a role for EMC in the co-translational biogenesis of GPCRs, defined the specific step at which EMC is needed, reconstituted its function in vitro, and used these insights to bypass EMC, thereby validating our conclusions in vivo.

Although we have focused on GPCRs, our mechanistic studies suggest that EMC is generally needed for initiating the accurate topogenesis of most membrane proteins containing an N_{exo} signal anchor. In support of this: a) all GPCR TM1 variants tested show some EMC dependency despite diverse sequences (Fig. 2.14 and Table 2.1), b) a non-GPCR (LepB) shows a minor dependence on EMC but can be made substantially EMC dependent by lengthening the TMD while maintaining overall TMD helicity (Fig. 2.17), c) EMC dependence does not act on a single sequence feature but appears to be necessary when a combination of biophysical features disrupt efficient insertion of N_{exo} signal anchors. Based on current prediction algorithms, 64% of the ~5000 human membrane proteins lack a signal peptide and therefore initiate topogenesis using a signal anchor. A large proportion of these adopt the N_{exo} topology and will therefore require EMC for optimal biogenesis. Thus, we propose that EMC is a widely conserved and extensively used insertion factor that operates in both co-translational and post-translational (Guna et al., 2017) membrane protein insertion at the eukaryotic ER.

Based on the available information, a working model for the role of EMC in signal anchor topogenesis can be proposed (Fig. 2.26). Extensive biochemical (Sakaguchi et al., 1987; Spiess and Lodish, 1986), structural (Hainzl et al., 2011; Janda et al., 2010; Keenan et al., 1998; Voorhees and Hegde, 2015), and whole-genome analysis (Chartron et al., 2016; Costa et al., 2018), suggest that a nascent signal anchor would be recognized by SRP and targeted to the ER membrane via the SRP receptor. At this stage, the signal anchor would be partially within or barely outside the ribosomal tunnel (Voorhees and Hegde, 2015), then dissociate from SRP and be located near the membrane. This is the critical time window in which EMC must act. At a later stage the accessibility to the nascent chain will likely be precluded by the tight interaction between Sec61 and the ribosome, biasing a nascent TMD towards Sec61. In the case of an N_{exo} signal anchor, this will result in either topological inversion or failure to insert altogether. The complete sufficiency and necessity of Sec61 for the insertion of N_{cyt} signal anchors [Fig. 2.19C and (Görlich and Rapoport, 1993)] supports the inference that topologically inverted β_1AR is inserted incorrectly via Sec61.

We favor a model in which N_{exo} substrates are inserted by EMC and the ribosome will then quickly dock on Sec61, positioning the second and remaining TMDs for insertion via the translocon (Fig. 2.24). This model is nicely supported by previous crosslinking studies demonstrating nascent N_{exo} signal anchors are near the translocon (Heinrich et al., 2000; High et al., 1993a). Our in vitro reconstitution of the biogenesis of multi-pass membrane proteins that depend on both EMC and the Sec61 complex, and the identification of specific steps where each is needed, now paves the way for mechanistic and structural dissection of this poorly understood process.

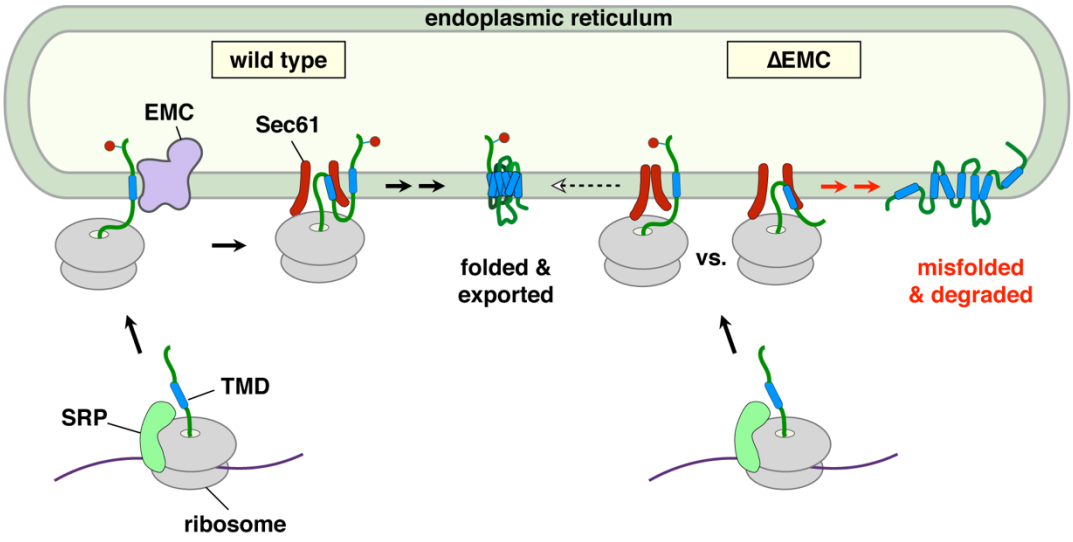


Fig. 2.26 - Working model for the roles of EMC and Sec61 complex in multipass membrane protein biogenesis. The left half of the diagram shows the normal situation (wild type), and the right half depicts the consequence of EMC deletion (ΔEMC). After targeting via SRP, the Nexo signal anchor is inserted via EMC. Downstream TMDs are inserted by Sec61. In the absence of EMC, the Nexo signal anchor of most of the nascent chains fails to insert in the correct topology, resulting in a misfolded and degraded protein. Depending on the substrate, some nascent chains are inserted appropriately by the Sec61 complex even in the absence of EMC, leading to a small population of correctly folded final protein.

Chapter III

An intramembrane chaperone facilitates multipass membrane protein biogenesis

The multipass IMP folding problem

Since the earliest studies on the denaturing and refolding of soluble proteins (Haber and Anfinsen, 1962), it has been well established that protein folds reside in free energy minima; adopting a final conformation that shields patches of polypeptide from energetically unfavorable interactions with the surrounding environment. However, the energetically preferred conformation is not always a functional product. For example, hydrophobic patches in cytosolic proteins can avoid the aqueous environment through productive folding or misfolding and aggregation (Pelham, 1986). A particularly vulnerable timeframe is during protein synthesis, when incomplete polypeptide is partially exposed to the cytosol but unable to interact with distal sequence features and fold into its final tertiary structure. Improper interactions are avoided through the action of molecular chaperones (Ellis, 1988; Pelham, 1986) which shield vulnerable segments of polypeptide, keeping them competent for folding. This process is well understood on a genome wide scale for soluble proteins and their associated co-translational chaperones have been thoroughly characterized (Blond-Elguindi et al., 1993; Döring et al., 2017; Stein et al., 2019).

Conversely, the assembly of multipass IMPs and multi-subunit membrane protein complexes is far more mysterious. An important distinction in intramembrane folding events is the inability of the lipid bilayer, as a solvent, to form hydrogen bonds (Popot and Engelman, 2000). Energetically, this places a large penalty on unpaired hydrogen donors/acceptors within a TMD, driving α-helix formation (Popot and Engelman, 1990) and simplifying the folding problem. While the ideal TMD is primarily hydrophobic on its surface (Engelman et al., 1986; Ulmschneider et al., 2005), the chemistry of purely aliphatic side chains is extremely limited and polar effects are necessary for forming the defined functional clefts and cavities observed in multipass IMPs (Engelman et al., 2003). However, polar properties within the membrane come with an energetic cost that must be offset to form an energetically stable protein (Chang and Bowie, 2014; Chen and Gouaux, 1999; Krishnamani et al., 2012; Lau and Bowie, 1997).

The large library of available membrane protein structures reveals how this is accomplished. TMDs are typically organized such that the hydrophilic regions face each other, presenting a mostly uninterrupted hydrophobic surface to the

surrounding lipid bilayer (Vinothkumar and Henderson, 2010). Thus, multi-spanning proteins often exploit polar and charged side chains for intramembrane interactions that facilitate TMD packing and protein folding (Harrington and Ben-Tal, 2009; Lear et al., 2000; Lu et al., 2018; Venkatakrishnan et al., 2013; Zhou et al., 2000). Similar to soluble proteins, nascent TMDs are most vulnerable to off pathway interactions during translation, when the full protein has not been synthesized. How semi-hydrophilic TMDs are temporarily stabilized within the membrane until their assembly with distal TMDs is not well understood.

It has been speculated that nascent TMDs might engage hypothetical chaperones within the lipid bilayer co-translationally (Shao and Hegde, 2011; Shurtleff et al., 2018). The physical proximity of a nascent TMD to various resident ER proteins has supported this idea (Andrews et al., 1991; Conti et al., 2015; McCormick et al., 2003; Meacock et al., 2002; Sadlish et al., 2005; Saksena et al., 2004), but these candidates are poorly studied. For example, the protein TRAM is found near signal sequences (Görlich et al., 1992b; High et al., 1993b) and some TMDs (Do et al., 1996; Heinrich et al., 2000), but no consequences of its depletion have been reported for either TMD insertion or membrane protein assembly. EMC has also been proposed to act as a chaperone for multi-spanning membrane proteins (Shurtleff et al., 2018), but this has been questioned by studies showing EMC is a TMD insertase (Guna et al., 2017) that aids in determining the first TMD's topology (Chitwood et al., 2018). Thus, ER-resident intramembrane chaperones that act on TMDs after their insertion have not been unambiguously identified.

An in vitro system for monitoring multipass membrane protein biogenesis.

After successful targeting and processing of a membrane targeting element (e.g. a TMD or signal peptide) the ribosome is firmly docked on the translocon and steric constraints bias newly translated polypeptide towards Sec61. Thus, TMDs are most vulnerable before Sec61 engagement and directly after integration. We sought to identify novel machineries acting to stabilize intermediates in the latter step, using incomplete multipass membrane proteins as a model. A truncated mRNA containing the first two TMDs of mammalian rhodopsin (Rho TM1+2) was translated in-vitro in the presence of human derived ER rough microsomes [Fig. 3.1A, similar to

(Meacock et al., 2002)]. Preliminary experiments established that tagged constructs coding for Rho TM1+2 were inserted in the appropriate orientation in hRMs isolated from HEK293 cells (3.1B). Insertion of both TMDs was confirmed by protease resistance followed by N-terminal immunoprecipitation and was unaffected by substitution with a single cysteine at position 53 or a variable N-terminal epitope tag (Fig. 3.1B, D). Accurate orientation could be inferred by both the size of the protected fragment and the addition of a glycan to an N-terminal acceptor site (Fig.3.1B). Consistent with this interpretation, direct comparison of a truncated RNC and a terminated version containing a stop codon confirmed complete and accurate insertion of the first 2 TMDs of rhodopsin using this system (Fig 3.1C, D).

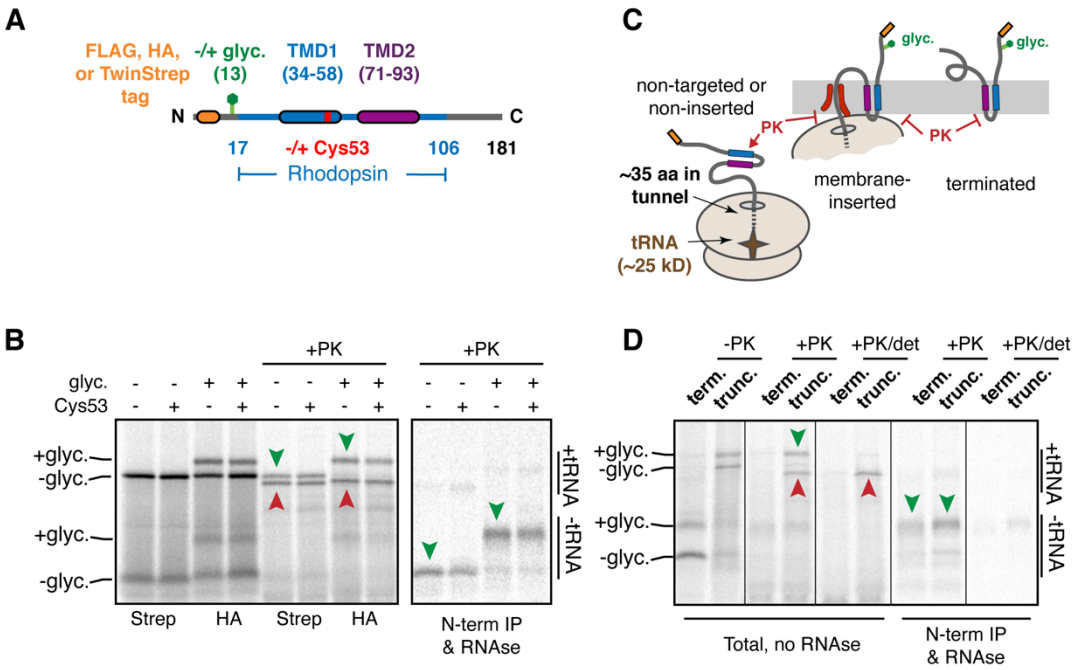


Figure 3.1 - Characterization of a multipass membrane protein intermediate. (A) Diagram of Rho TM1+2 constructs used throughout this chapter. Variations on this construct include different N-terminal epitope tags, the presence or absence of a glycosylation site near the N-terminus, the presence or absence of a cysteine within TMD1, various mutations within TMD1, and the presence and identity of TMD2. All of the constructs were tested either by protease protection or glycosylation to verify that no appreciable differences were observed in their insertion efficiencies. Note that although the exact amino acid numbering varies depending on the N-terminal tag, the numbering system corresponding to the FLAG-tagged version is used throughout. Thus, the 146mer refers to a truncation at the 146th codon in the numbering scheme indicated, even in constructs containing a different tag. **(B)** Representative example of insertion assays on two different tagged versions of Rho TM1+2. The TwinStrep tagged version (Strep) lacking a glycosylation site was compared to an HA tagged

version containing a glycosylation site. Identical constructs containing either the wild-type Rho TM1 sequence or a point mutant (F53C) were tested in parallel to confirm no insertion defects result from insertion of a cysteine in TM1 (used for BMH-mediated crosslinking in later experiments). In this experiment, ribosome nascent chain complexes (RNCs) of 181 amino acids were produced by translation in the presence of ER-derived rough microsomes (RMs) after which the microsomes were isolated and resuspended. Aliquots of the reactions were left untreated or digested with proteinase K (PK) and analyzed directly by SDS-PAGE (left panel). Green arrow heads represent the fully inserted and PK-protected population and red arrowheads denote the non-inserted and proteolytically cleaved products. The cleaved product contains the region of polypeptide protected by the ribosomal tunnel and the attached tRNA. Aliquots of the PK-digested sample were treated with EDTA and RNase to release the polypeptide from the ribosome and digest tRNA and immunoprecipitated (IP) via the N-terminal tag. Only the fully inserted products are recovered by IP (green arrowheads). **(C)** Diagram representing the interpretation of the experiments in panels B and D. The relatively short cytosolic loop between TMD1 and TMD2 is not accessible to PK digestion either as an RNC or a terminated product. **(D)** Comparison of the topology of truncated RNCs and terminated Rho TM1+2. In this experiment, the FLAG-tagged Rho TM1+2 containing a glycosylation site with (term.) or without (trunc.) a stop codon was translated in the presence of RM after which the RMs were isolated by centrifugation. Aliquots of the isolated RMs were analyzed directly (-PK), after PK digestion (+PK), or after PK digestion in the presence of detergent (+PK/det). Where indicated, the +PK and +PK/det samples were released from the attached tRNA and immunoprecipitated via the N-terminal tag. Note comparable glycosylation near the N-terminus and complete protection from PK for both the truncated and terminated products. All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 25% of the total reaction.

Detection and characterization of an intramembrane interacting factor

With the above tools and assays in hand, we took advantage of staged integration intermediates and the cysteine specific crosslinking reagent, bismaleimidohexane (BMH), to monitor the molecular environment of TM1 during different stages of insertion (Fig 3.2). Stalled translation complexes containing different lengths of nascent polypeptide were crosslinked to adjacent proteins via a single cysteine introduced into TMD1 (Fig. 3.2 upper cartoon). Consistent with previous reports (Keenan et al., 2001; Rapoport, 2007), early intermediates (lanes 1-3) crosslinked to well-characterized ribosomal and ribosome-associated proteins including the SRP54 subunit of the targeting factor SRP (Keenan et al., 2001) and the Sec61α subunit of the translocation channel (Rapoport, 2007). Thus, this assay provided a faithful readout for the position of a TMD and the relative timing of engagement of several established interacting partners during insertion. Intermediates where TMD2 was emerging from the ribosome (lanes 4-6) showed diminished TMD1-Sec61α crosslinks and the appearance of crosslinks to a ~10 kD protein which we presume is

identical to a previously observed crosslinking partner dubbed PAT10 for "protein associated with the ER translocon of 10 kD" (Meacock et al., 2002). At all subsequent lengths, the PAT10 crosslink persisted at approximately the same efficiency while the Sec61α crosslink remained low.

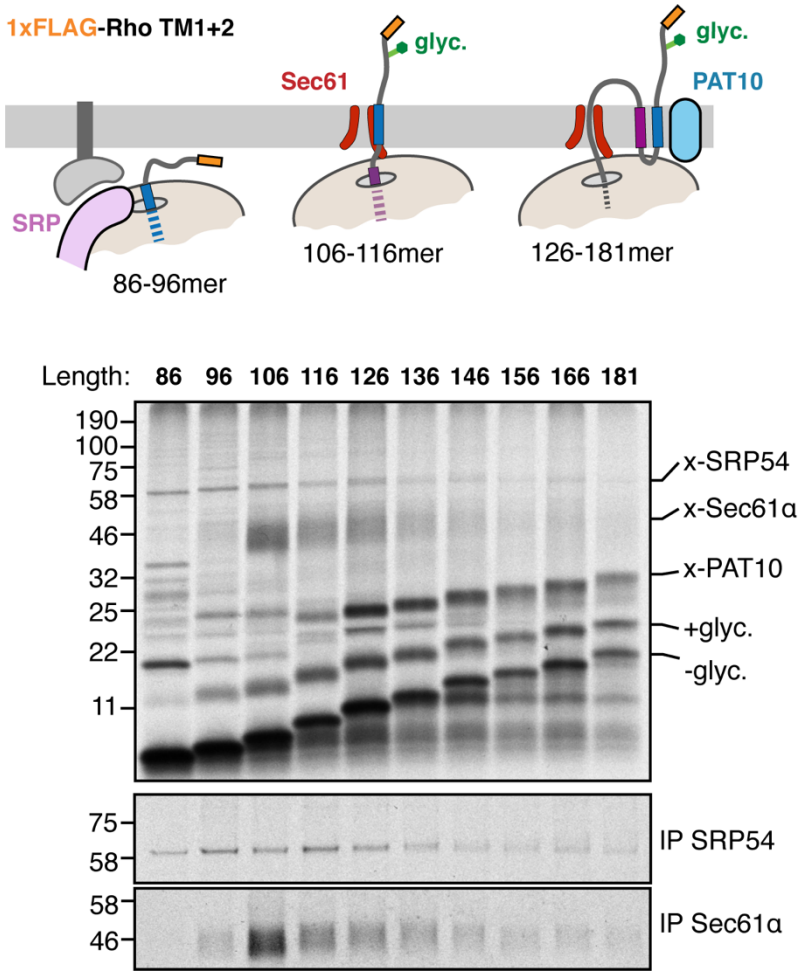


Figure 3.2- A protein engages nascent membrane proteins. Cysteine-based crosslinking of ^{35}S-labeled ribosome nascent chain complexes (RNCs) representing intermediates during targeting and insertion of a rhodopsin-based construct (see upper diagram) encoding the first two transmembrane domains (see Fig. 3.1A). Nascent chains ending at the indicated amino acid were produced by in vitro translation containing ER-derived rough microsomes (RMs) from HEK293 cells, treated with bismaleimidohexane (BMH; a cysteine-reactive crosslinker), and subject to denaturing immunoprecipitation (IP). The upper gel shows the translation products and all of their crosslinks as visualized by autoradiography of IPs via the nascent chain's N-terminal FLAG tag. Indicated on the right are the non-glycosylated (-glyc.) and glycosylated (+glyc) translation products and the crosslinks to PAT10, Sec61α, and SRP54 (verifed by IPs; lower gels). All of the RNCs contain attached tRNA which is removed by RNAse digestion prior to analysis on SDS-PAGE (see Fig. 3.1). All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 25% of the total reaction.

Importantly, the observed crosslinking adducts were highly specific to the addition of BMH and observed with the tRNA associated species, consistent with co-translational association (Fig. 3.3). These experiments also allow for the independent analysis of TMD insertion via the N-terminal glycosylation site, early intermediates show relatively little glycosylation (pre-insertion) while later intermediates demonstrate an increase in insertion efficiency as determined by the ratio of glycosylated to unglycosylated products (Fig. 3.3). One technical point is worth highlighting; the observed decrease in intensity for the PAT10 adduct at longer nascent chain lengths does not reflect a decrease in PAT10 association, but instead reflects decreased translation efficiency (i.e. total product, see left panel). A key metric is the ratio of intensity between the inserted product (glyc.) and the PAT10 crosslinked product, which objectively remains relatively constant throughout extension of the nascent chain (Fig. 3.2 and 3.3).

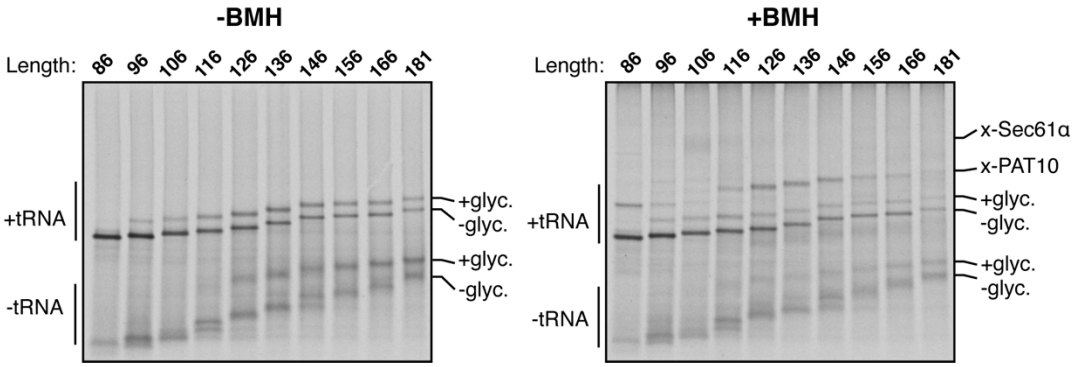


Figure 3.3 - Characterization of insertion and crosslinking to Rho TM1+2 RNCs. ^{35}S-labeled FLAG Rho TM1+2 ribosome nascent chains (RNCs) of varying lengths (as described in Figure 3.2) were generated by in-vitro translation in the presence of RMs. Membranes were isolated by centrifugation through a sucrose cushion and resuspended in physiological salt buffer (PSB). An equivalent amount of each translation reaction was taken before (-BMH) and after (+BMH) the addition of BMH for analysis by SDS-PAGE. The tRNA-linked nascent chains and free nascent chains are indicated. The free nascent chains arise from partial hydrolysis of the tRNA during electrophoresis under moderately basic conditions. Glycosylation is first observed at the 96mer length, which is 38 amino acids downstream of the end of TMD1. This matches the length of the ribosome tunnel, and indicates that membrane insertion and glycosylation occurs only after the full TMD is exposed outside the ribosome. Crosslinks to Sec61α are most prominent for the 106mer. Crosslinks to PAT10 are most prominent from the 126mer onwards, after the Sec61α crosslinks diminish. Note that all of the crosslinked adducts are seen to the tRNA-attached nascent chain, verifying that they are co-translational. The Sec61α crosslink and others are not as visible when total translation products are analyzed, which is why we typically immunoprecipitate the

sample via the nascent chain (e.g., in Fig. 3.2). Furthermore, we usually digest the samples with RNAse after the experiment but before SDS-PAGE to remove the tRNA, thereby avoiding the heterogeneity.

Modification of the sequence downstream of TMD1 allowed us to manipulate the relative timing of TMD1 migration into the lipid bilayer and correspondingly monitor PAT10 engagement (Fig. 3.4A). Initial direct comparisons of Rho TM1+2 variants, at a single nascent chain length, displayed no significant differences in translation or insertion efficiency (Fig. 3.4B). Replacement of TMD2 with an efficiently inserted TMD from asialoglycoprotein receptor (ASGR) displaced TMD1 from Sec61α more efficiently than native TMD2 (Fig. 3.4C). By contrast, replacement of TMD2 with a flexible soluble segment of polypeptide delayed TMD1 movement away from the Sec61 complex (Fig. 3.4D). In each case however, the appearance of PAT10 crosslinks coincided with the diminishment of Sec61α crosslinks. Thus, PAT10 engages TMD1 upon its departure from the Sec61 complex, placing its functional role at step downstream of TMD insertion. This result is consistent with previous observations in Figures 3.2-3.3 demonstrating a persistent PAT10 crosslink.

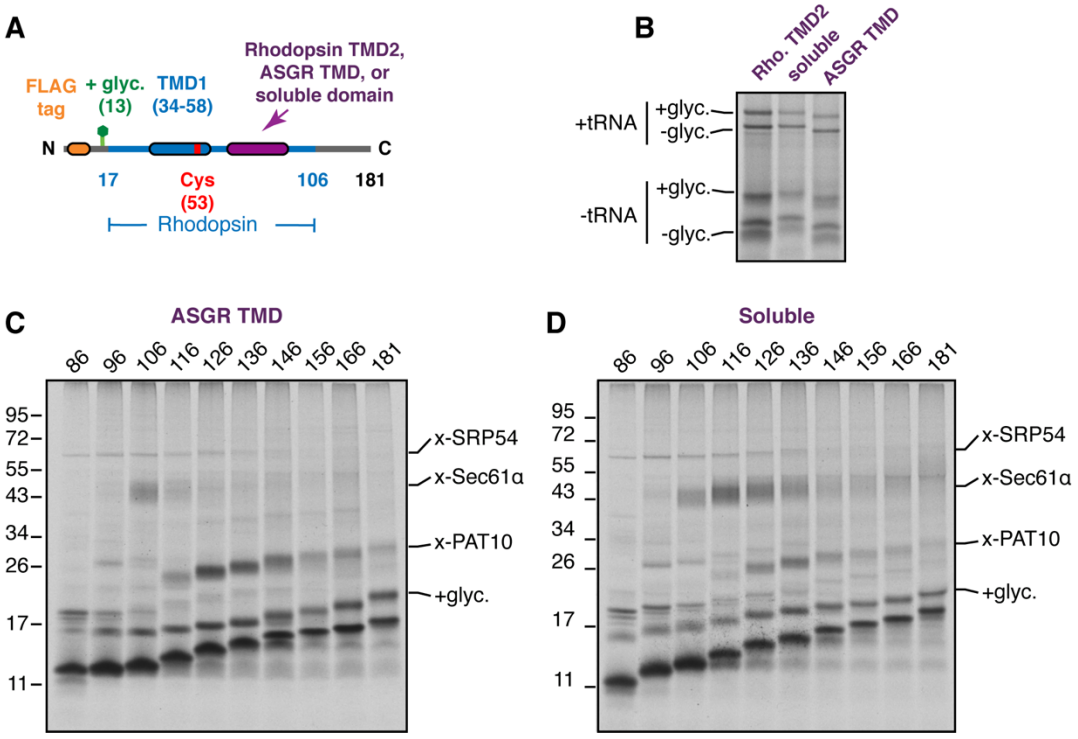


Figure 3.4 - PAT10 engagement occurs after the TMD departs from Sec61α. (A) Diagram of the three matched constructs used to assess timing of PAT10 engagement via BMH crosslinking to the single cysteine located in Rho TMD1. A parent Rho TM1+2 construct (Figure 3.2) was modified by replacement of TMD2 with either the TMD of asialglycoprotein receptor (ASGR1) or a soluble domain. **(B)** Radiolabeled 146mer RNCs encoding the 3 different construct variants shown in panel A were generated by in-vitro translation in the presence of RMs and analyzed by SDS-PAGE to visualize total translation products. No appreciable difference in insertion efficiency was observed between constructs as monitored by glycosylation efficiency. SDS-PAGE causes a small amount of hydrolysis of the tRNA-peptidyl bond and both the tRNA associated (+tRNA) and the non-tRNA associated (-tRNA) products can be observed. **(C, D)** Analysis of Rho TM1-ASGR1 and Rho TM1-soluble exactly as in Fig. 3.2. Notice the abrupt decrease in crosslinking efficiency of TMD1 to Sec61α observed in panel C. This suggests that the hydrophobic TMD of ASGR1 displaces TMD1 from Sec61α. By contrast, TMD1 crosslinking to Sec61α is observed over a much wider range of lengths when it is not followed by another TMD, suggesting that it remains near the Sec61 complex. In both cases however, the appearance of crosslinks to PAT10 correlate with the decrease in Sec61α crosslinks. Thus, PAT10 engagement occurs coincident with TMD1 moving away from Sec61α.

Substrate crosslinking to PAT10 did not depend on glycosylation (Fig. 3.5A), and crosslinks were also observed from other cysteine positions within TMD1 (Fig. 3.5B-C), consistent with earlier observations (Meacock et al., 2002). Additionally, a very weak crosslink to the β subunit of Sec61 was observed only after substantial enrichment via immunoprecipitation using a Sec61β specific antibody (Fig. 3.5A). A noteworthy point is the presence of a single cysteine in the unstructured cytosolic domain of Sec61β, consistent with crosslinks to the non-inserted (unglycosylated) Rho TM1+2 and the substantially reduced crosslinking efficiency in comparison to factors that directly engage a TMD (e.g. SRP54, Sec61α). Conversely, PAT10 crosslinking to inserted (i.e., glycosylated) TMD1 was over extremely efficient (Fig. 3.2 and Fig. 3.5A) indicating that a high proportion of inserted nascent chains sample this factor. PAT10 appears to be both conserved and widely expressed, as the interaction is preserved in a system using ER rough microsomes derived from canine pancreas (Fig. 3.5C). In combination with the high degree of substrate engagement after insertion, this argues for a general factor that engages TMDs in the membrane.

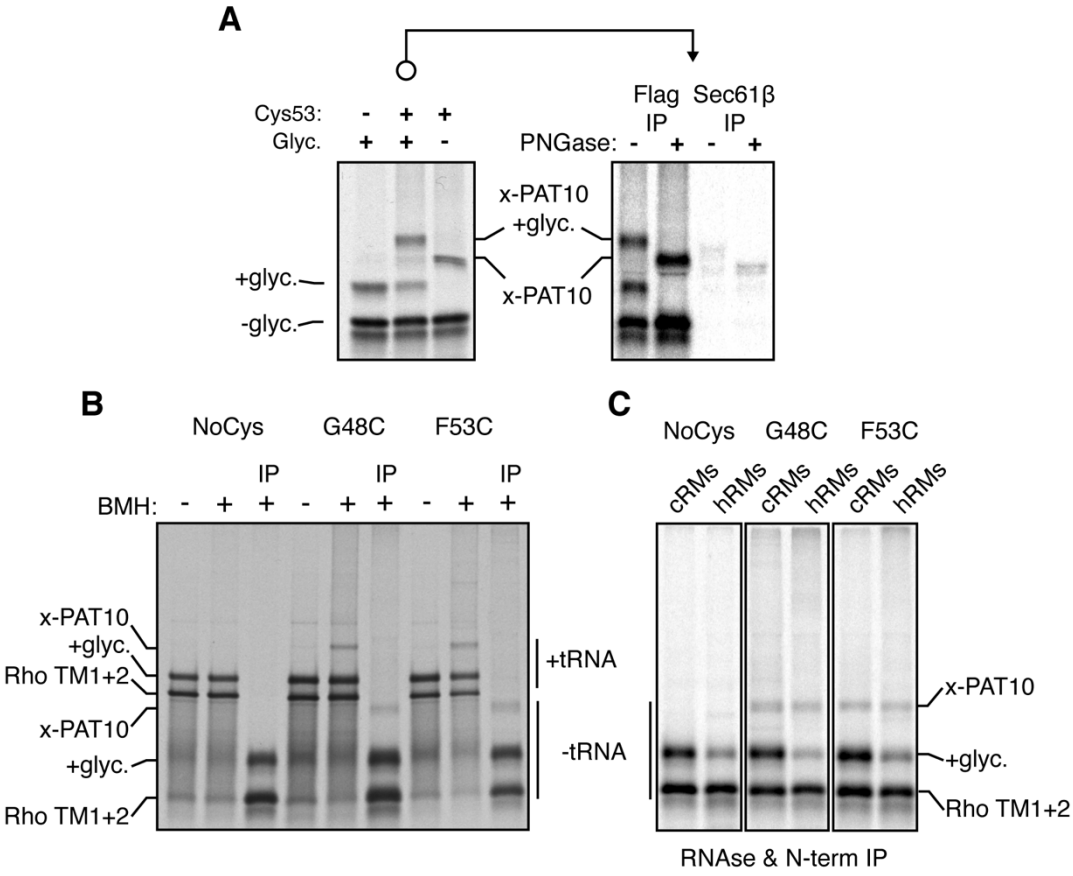


Figure 3.5 - PAT10 engages inserted TMDs independent of glycosylation. **(A)** Radiolabeled 146mer RNCs of FLAG-tagged Rho TM1+2 were produced by in vitro translation in the presence of RMs, then treated with BMH. The construct either contained or did not contain a cysteine at position 53 (Cys53) and/or a glycosylation site near the N-terminus (Glyc). The crosslinking reactions were digested with RNAse and either analyzed directly (left panel), or the indicated reaction was subjected to immunoprecipitation (IP) via the FLAG tag or Sec61β under denaturing conditions. The IP samples were either left untreated or digested with PNGase F to remove the N-linked glycans. Equivalent amounts were loaded in each lane. Note that Sec61β is not an appreciable crosslinking partner of these RNCs, and to the extent a crosslink is observed, it migrates slightly faster than the PAT10 crosslink. **(B)** The indicated Rho TM1+2 variants were translated in vitro in the presence of RMs and treated with BMH as indicated. In this experiment, the crosslinking was done directly on total translation reactions, not after isolation of the microsome fraction. Instead, translation reactions were diluted 5-fold with buffer to dilute the reduced glutathione and reduce quenching of BMH. An aliquot of each reaction was analyzed directly by SDS-PAGE. Cross-linking efficiency is reduced compared to other experiments because membranes were not isolated by centrifugation through a sucrose cushion to remove reduced glutathione from translation extract. One aliquot of the BMH-treated translation reactions was treated with RNase A and EDTA, denatured, and IPed via the N-terminal FLAG tag (IP). **(C)** As in panel B except Rho TM1+2 variants were generated in the presence of RMs derived from either canine pancreas (cRM) or HEK293 cells (hRM). In this experiment, the microsomes were isolated by centrifugation through a sucrose cushion prior to BMH crosslinking (note higher crosslinking efficiency). While the PAT10 crosslink is seen in both cRM and hRM, crosslinking efficiency of the inserted (glycosylated) product is significantly lower in cRMs, which is one reason we

used HEK293-derived RMs for most of the experiments. All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 25% of the total reaction.

The extremely efficient interaction observed between PAT10 and a nascent TMD (as judged by crosslinking efficiency) argued strongly for a coordinated interaction with newly translated membrane proteins at the translocon. We reasoned directed action and engagement of TMDs was unlikely to be coordinated by a single ~10kDa membrane protein. In support of this, the natively solubilized ribosome-released TMD1-PAT10 complex migrated at an apparent molecular weight of >100 kD by sucrose gradient sedimentation (Fig. 3.6). This is far larger than would be expected if the interaction were between a PAT10 molecule in isolation (~10kDa) and a nascent chain (~14kDa). PAT10 is therefore part of a larger assembly **(the PAT complex)** which is adjacent to a TMD as it releases from the Sec61 translocon. Because efficient maleimide-mediated crosslinking requires an aqueous environment (Lee and Samuels, 1964), the PAT complex seems to provide a membrane-embedded partially hydrophilic environment where TMD1 binds.

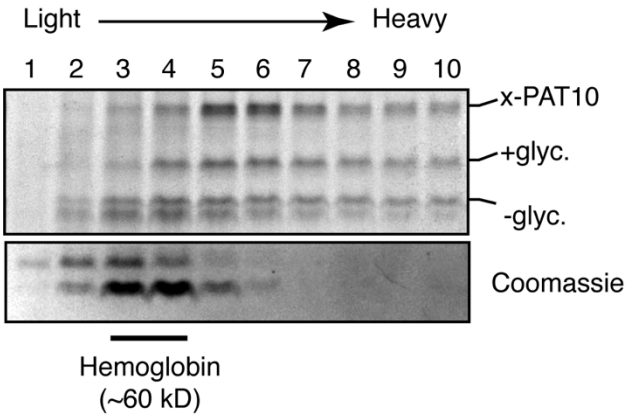


Figure 3.6 – PAT10 is part of a larger molecular weight complex (the PAT complex). Sucrose gradient separation of the ^{35}S-labeled membrane-targeted 146mer RNC after BMH crosslinking, native solubilization, and release from the ribosome by RNAse digestion. The migration of endogenous hemoglobin (~60 kD) from the translation extract was visualized by Coomassie staining of the gel.

The PAT complex is composed of CCDC47 and Asterix

To identify components of the PAT complex, we solubilized the crosslinking reactions under non-denaturing conditions, released the nascent chain from the

ribosome, affinity purified via the FLAG tag on the nascent chain, and identified the co-purifying proteins by quantitative mass spectrometry (Fig. 3.7). Six proteins were enriched more than 2-fold relative to a control substrate containing a Strep tag: Sec61α and Sec61β, the lectins Calnexin and Galectin-7, signal peptide peptidase (SPP), and CCDC47. Of note, none of these proteins was a ~10 kD candidate for PAT10.

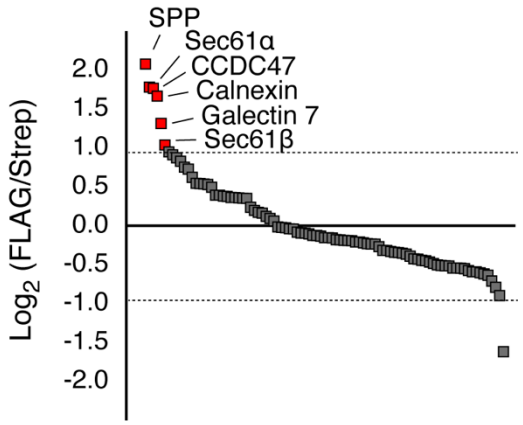


Figure 3.7 – Identification of potential PAT complex members. BMH-crosslinked 146mer RNCs containing a FLAG or Strep tag were released from the ribosome by RNAse digestion, subjected to native FLAG IPs, and analyzed by quantitative mass spectrometry. Proteins enriched 2-fold or more in the FLAG-tagged RNCs are indicated. Note, the value plotted on the y-axis is the enrichment of peptides recovered in the experimental (FLAG) vs control (Strep tag) and therefore does not reflect abundance of products but rather their signal relative to background recovery of proteins during the IPs.

Sec61α and Sec61β presumably were recovered due to their direct crosslinking to the nascent chain (Fig. 3.2 and Fig. 3.5B). Preliminary analysis of the complex bound to Rho TM1+2, with the glycosylation acceptor site removed, demonstrated a substantial decrease in molecular weight by sucrose gradient fractionation (Fig. 3.8A). Consistent with this observation, native solubilization of crosslinked substrate followed by immunoprecipitation demonstrated a glycan-dependent association with Calnexin (Fig. 3.8B). Therefore, lumenal acting chaperones (e.g. lectins) can simultaneously engage substrate with PAT-10, but these interactions are uncoupled. Similarly, SPP demonstrated direct crosslinks to Rho TM1+2 (Fig. 3.8C), as previously reported (Crawshaw et al., 2004; Schrul et al., 2010). However, these were comparatively weak and not observed unless enriched using immunoprecipitations against SPP (Fig. 3.8C). Furthermore, under native IP conditions SPP did not

enrich for the PAT10 associated complex. Rather a subtle but significant enrichment for the non-inserted product was observed (see unglycosylated substrate in SPP IP).

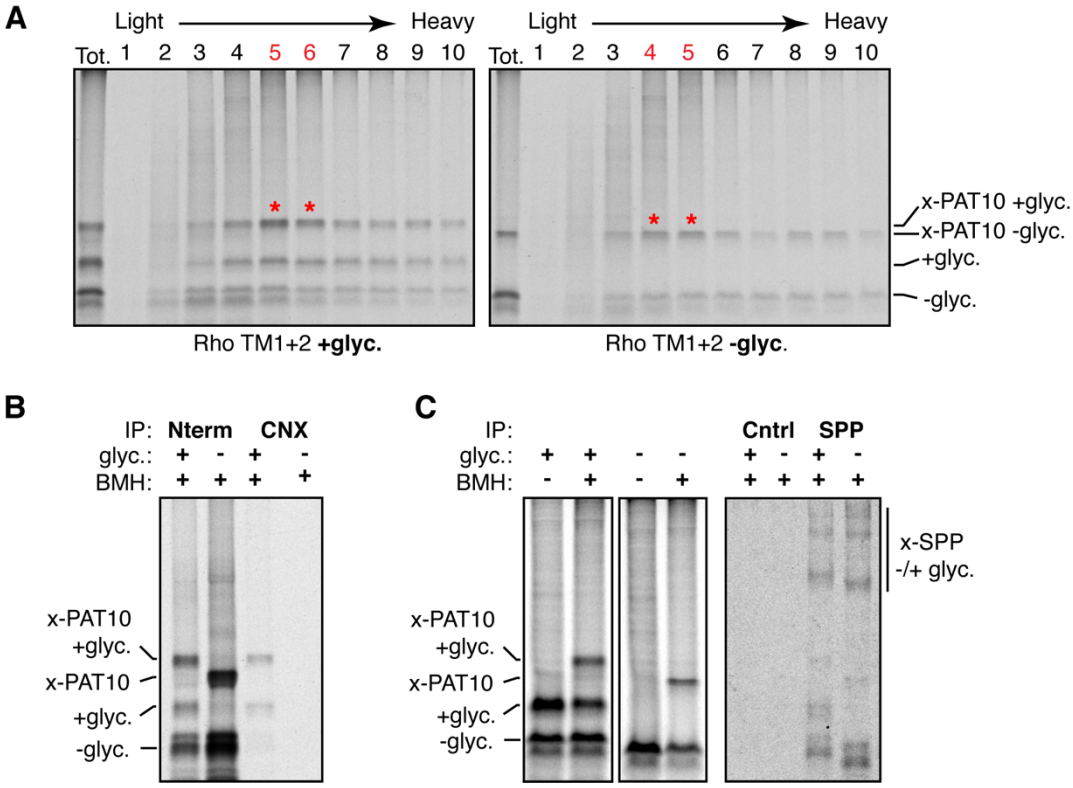


Figure 3.8 – Characterization of non-PAT complex TMD interactors. (A) 146mer RNCs containing a FLAG tag were generated by in vitro translation in the presence of RMs. Membranes were isolated by centrifugation through a sucrose cushion, subjected to BMH crosslinking, released from ribosomes and treated with RNase A followed by native solubilization and fractionation on a 5-25% sucrose gradient before analysis by SDS-PAGE and autoradiography. Red asterisks denote peak fractions containing the PAT complex as detected by the PAT10 crosslinking product. The PAT complex crosslinked to unglycosylated Rho TM1+2 migrates slightly smaller on the gradient than glycosylated Rho TM1+2, likely the result of CNX (~90 kD) no longer associated with the nascent chain. **(B)** Radiolabeled 146mer RNCs of FLAG tagged Rho TM1+2 containing (-) or lacking a glycosylation site (+) were generated by in vitro translation with RM, crosslinked with BMH, then immunoprecipitated under native conditions via the N-terminal FLAG tag on the substrate (Nterm) or with an antibody recognizing the C-terminus of calnexin (CNX). Only the glycosylated substrate is recovered with CNX, consistent with its binding via the glycan. **(C)** Aliquots of samples in panel A containing or lacking a glycosylation site were generated then analyzed by SDS-PAGE. One aliquot of the BMH treated reactions were solubilized under native conditions and IPed with either an antibody raised against an HA epitope tag (Cntrl) for a specificity control, or an antibody against signal peptide peptidase (SPP). As previously reported, direct crosslinks to SPP are observed as two distinct adducts seen on very long exposures, but native IPs do not enrich for a PAT10 engaged substrate. All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 25% of the total reaction.

For the remainder of our efforts, we focused on CCDC47 because unlike the others, the PAT10 crosslink was strongly enriched in native IPs using anti-CCDC47 antibodies (Fig. 3.9A). Direct crosslinks to CCDC47 were never observed, however inspection of the CCDC47 native TMD sequence indicated no cysteines so a direct interaction with substrate cannot be ruled out. Even in the absence of crosslinking, CCDC47 selectively enriched the inserted population of nascent chains (Fig. 3.9B). This is in contrast to SPP which fails to pull out the PAT10 crosslinked substrate, or any substrate in the absence of BMH, when the two are directly compared by native immunoprecipitation (Fig. 3.9A-B). Direct crosslinks to SPP are only visible after long exposure times and using large scale reactions that are then IPed with SPP (Fig. 3.8C). This interaction was specific because recovery of the nascent chain was completely lost when microsomes prepared from *CCDC47* knockout cells were used (Fig. 3.9A), or when the sample was denatured prior to immunoprecipitation.

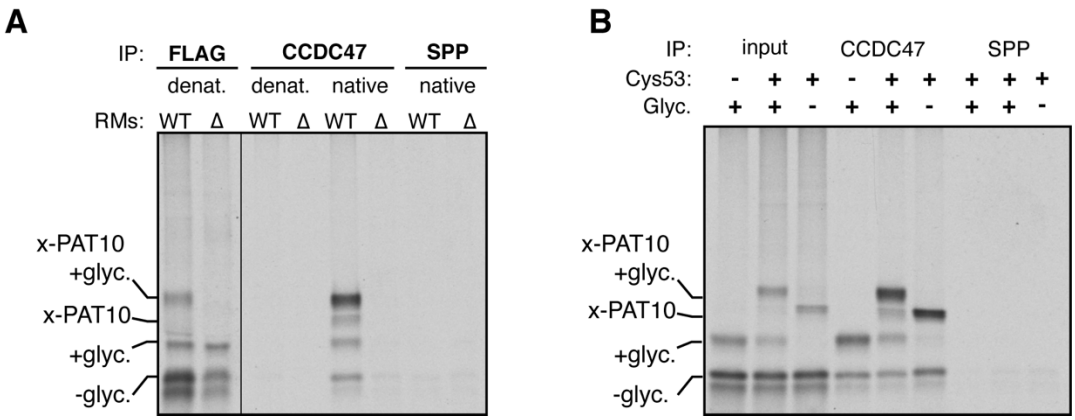


Figure. 3.9 – CCDC47 is the primary interacting partner of the TMD-PAT10 complex. (A) 146mer RNCs targeted to RMs prepared from WT or ΔCCDC47 (Δ) cells were treated with BMH, released from the ribosome with RNAse, and IPed under denaturing (denat.) or native conditions with the antibodies indicated. **(B)** 146mer RNCs containing or lacking either a glycosylation site (Glyc.) or cysteine at position 53 (Cys53) were crosslinked with BMH and analyzed directly (input) or after native IP using antibodies against CCDC47 or signal peptide peptidase (SPP). Nascent chains were released from the ribosome with RNAse A before IP. All immunoprecipitations were loaded such that the input represents 10% of the total reaction and the IP represents 25% of the total reaction.

Importantly, CRISPR mediated knockout of CCDC47 completely abrogated protein expression in HEK cells (Fig. 3.10A). Strikingly, TMD1 crosslinking to PAT10 was strongly reduced in microsomes lacking CCDC47 (Fig. 3.10B). Together with the data demonstrating direct interaction with a PAT10 engaged substrate (Fig. 3.9), this firmly establishes CCDC47 as a previously unknown interactor of nascent membrane proteins and a stable subunit of the PAT complex.

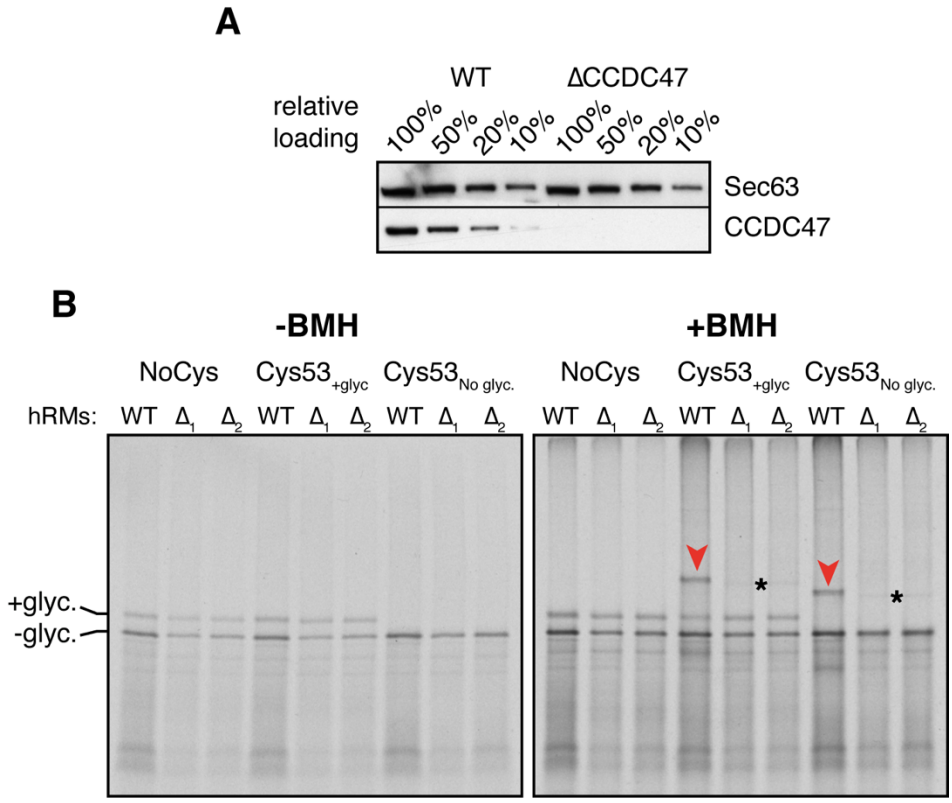


Figure 3.10 – Substrate crosslinking to PAT10 is dependent on CCDC47. (A) Western blot of RMs prepared from wild type (WT) or CCDC47 knock-out (ΔCCDC47) HEK293 cells comparing CCDC47 levels relative to Sec63, another ER resident protein. (B) The insertion and BMH mediated cross-linking for 146mer RNCs of the parent Rho TM1+2 construct or versions lacking a cysteine (NoCys) or lacking a glycosylation site (Cys53No Glyc.). The radiolabeled RNCs were produced by in vitro translation in the presence of RMs isolated from wild type (WT) or two different CCDC47 KO cell lines (Δ1 and Δ2) generated from two different guide RNAs. Aliquots of the reaction before (-BMH) and after (+BMH) addition BMH were analyzed by SDS-PAGE. No appreciable difference in insertion efficiency was observed in KO microsomes for Rho TM1+2 as monitored by glycosylation efficiency. Red arrowheads indicate the PAT10 crosslink which is lost upon CCDC47 KO. The faint crosslinked adduct observed in the KO samples (black asterisks) migrates slightly faster on the gel and likely represents weak Rho TM1+2 crosslinks to Sec61β (see Extended Data Fig. 4).

A stable association between CCDC47 and PAT10 provided a new molecular handle to isolate and identify PAT10 from natively solubilized hRMs. Affinity columns made with two different anti-CCDC47 antibodies each recovered CCDC47 and a prominent ~10 kD protein (Fig. 3.11A). While mass spectrometry of tryptic digests failed to identify this protein (explaining why it was not identified in the initial mass spectrometry experiment above), digestion with other proteases produced peptides that matched Asterix (also called WDR83OS). No other stoichiometric and specific interaction partners were observed in the purifications or identified by mass spectrometry. Additionally, Asterix contained putative transmembrane domains, was 12kDa in size, and contains cysteines within its predicted TMDs (See Fig. 3.12). Further analysis by direct immunoprecipitation of Asterix (Fig. 3.11B), and its immunodepletion during CCDC47 pull-downs (Fig. 3.11C), confirmed Asterix directly interacts with newly inserted Rho TM1+2 and forms a stable complex with CCDC47.

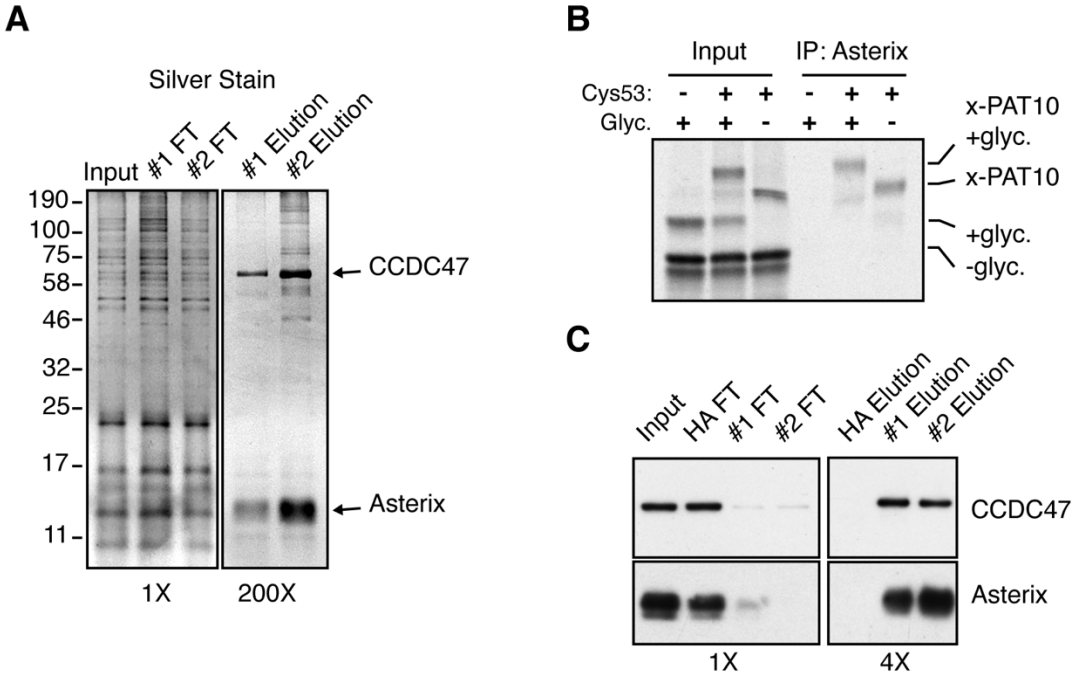


Figure 3.11 – Asterix and CCDC47 form the PAT complex. **(A)** Affinity purification of CCDC47 from natively solubilized RMs using two unrelated CCDC47 antibodies (#1 and #2). The elution samples represent ~200-fold more than the input and flow-through (FT) samples. **(B)** ^{35}S-radiolabeled 146mer RNCs of Rho TM1+2 containing or lacking Cys53 or a glycosylation site were targeted to RMs, treated with BMH, digested with RNAse, and analyzed directly (input) or after denaturing IP with anti-Asterix antibodies. **(C)** Affinity purification via CCDC47 as in panel A, but with a negative control using anti-HA antibodies. Asterix and CCDC47 were detected by immunoblot. The elution samples represent 4-fold more than the input and FT samples. Unless

otherwise indicated in the figure panel, all immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 25% of the total reaction.

The PAT complex

CCDC47 and Asterix are poorly studied but are widely conserved across all eukaryotes (Figure 3.12A-B) and are widely expressed in all tissues. Furthermore, loss of CCDC47 in mice is embryonic lethal with a wide range of developmental problems (Yamamoto et al., 2014) and mutations in humans causes numerous developmental phenotypes with early death (Morimoto et al., 2018). Cells lacking either subunit of the PAT complex show reduced fitness across multiple mammalian cell lines (Wang et al., 2015) but are viable, show activated ER stress responses, and have several otherwise unrelated membrane-associated phenotypes (Yamamoto et al., 2014). While observations on the consequences of Asterix depletion at the level of an organism are absent, CCDC47 and Asterix form an obligate complex because knockout of either protein results in substantial loss of the other (Fig. 3.12C), although the exact stoichiometry remains to be determined. Consistent with a role in protein homeostasis, genetic knockout of the PAT-complex results in elevated levels of the ER chaperones BiP and PDI (Fig. 3.12C), however it does not appreciably affect the expression levels of any previously established biogenesis proteins (Fig. 3.12C). At the expression level, the PAT complex is distinct from EMC as both are unaffected by loss of the other, although both seem to induce an unfolded protein response as judged by similar increases in BiP and PDI.

Further analysis of the PAT complex predicts CCDC47 to be an ER-resident single-pass membrane protein with a well-conserved cytosolic domain (Fig. 3.12). Topology analysis of Asterix by in vitro translation and protease protection assays suggest that it is a three-TMD membrane protein whose N-terminus faces the cytosol and C-terminus faces the ER lumen (Fig. 3.13A-B). Reintroduction of Asterix variants in ΔAsterix HEK cells and PEG-Maleimide labelling of single cysteines, placed as topological identifiers, support a 3-TMD N_{cyt} topology (Fig. 3.13C-D). Of note, the three putative TMDs are only ~15 amino acids each, populated by numerous moderately hydrophilic amino acids, and contain multiple cysteines that explain its efficient crosslinking to nascent TMDs via BMH (Fig. 3.12B). These features, when

considered in the context of the phenotypic consequences of PAT complex depletion described in other studies, and PAT complex engagement of a TMD released from the translocon, strongly argue for a role in membrane protein biogenesis.

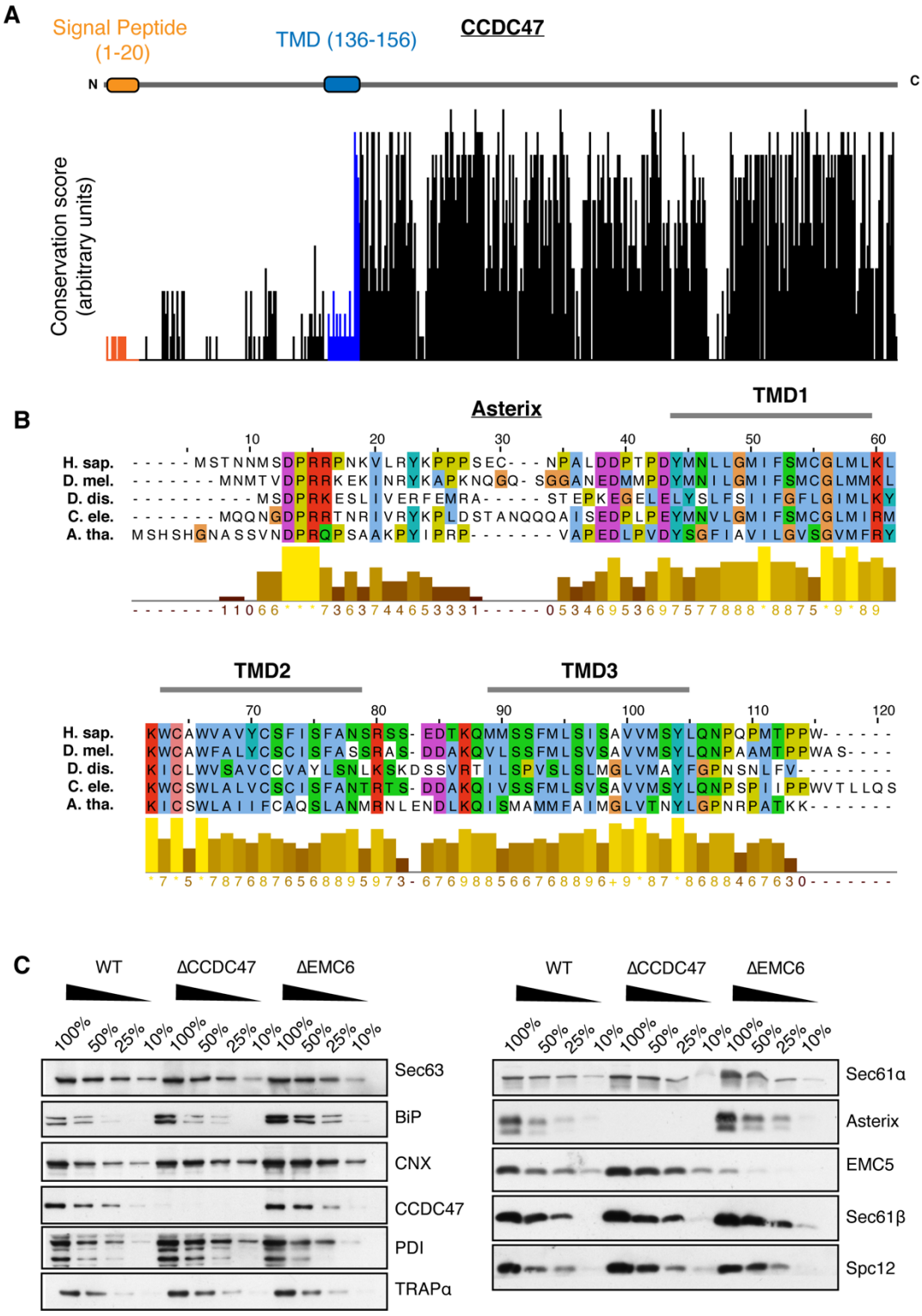


Figure 3.12 – PAT10 and Asterix are widely conserved and associated with protein homeostasis. **(A)** Conservation score plotted for each amino acid of human CCDC47. Annotated sequences for CCDC47 homologs were downloaded from the OMA orthology database and aligned using the Clustal W alignment plugin in Jalview. The conservation scores were assigned based on physio-chemical properties of amino acids, with identities scoring highest. **(B)** Alignment of Asterix homologs for five divergent species with a bar chart representing conservation scores of each amino acid. TMDs were based on the annotation in Uniprot, except that the very long single TMD from amino acids 44 to 78 was deduced to represent two closely spaced TMDs (see Fig 3.13). Note the high abundance of hydrophilic amino acids within the TMD regions and also the presence of multiple cysteines, explaining the robust crosslinking efficiency observed with BMH. **(C)** ER rough microsomes were isolated from WT, ΔCCDC47 and ΔEMC6 HEK293 cells and normalized to an absorbance of 75 at 280nm. Serial dilutions of each sample were analyzed by SDS-PAGE and immunoblotting for the indicated antigens. Note that BiP levels are elevated in both knockout cell lines consistent with an activated UPR caused by altered ER homeostasis.

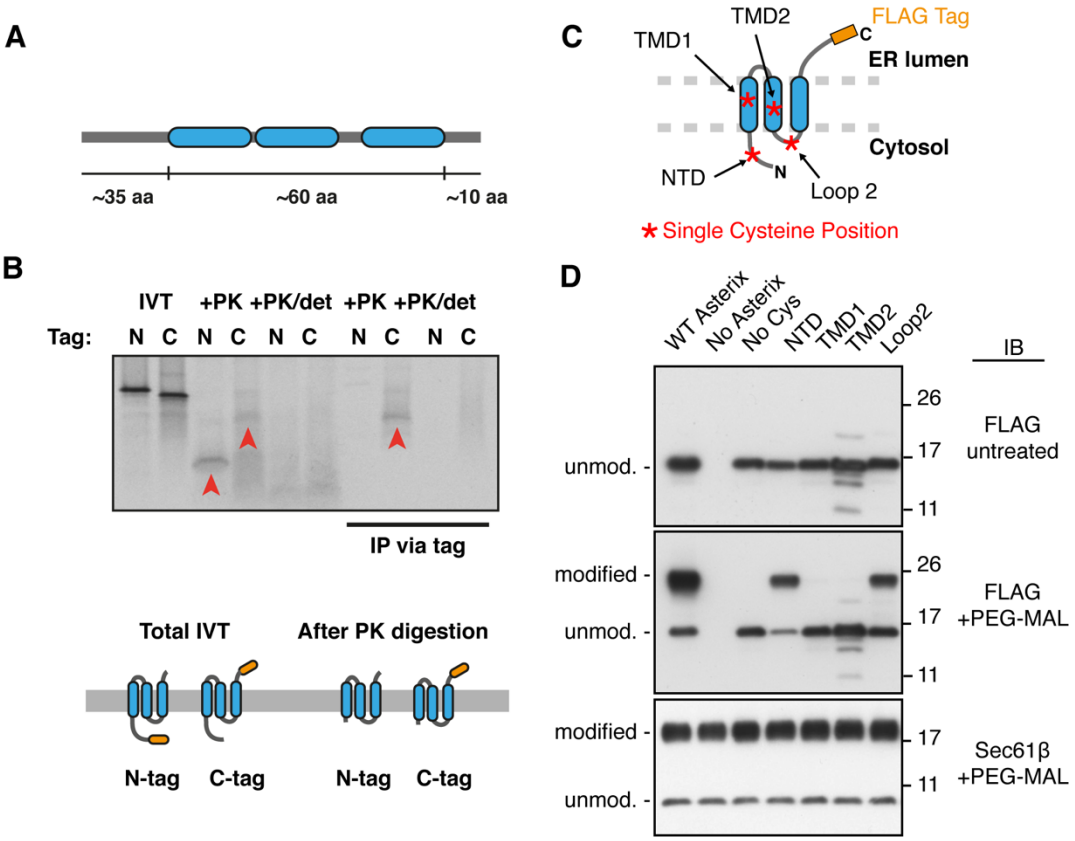


Figure 3.13 – The topology of Asterix. (A) Schematic of the relative size of Asterix and its topologically defined domains. **(B)** Two matched human Asterix constructs containing either an N- or C-terminal FLAG tag were generated by in vitro translation in the presence of RMs. One aliquot of the reaction was set aside for analysis by SDS-PAGE of total translation products (IVT). The remainder of the reaction was treated with proteinase K (+PK) without or with detergent (det). These protease-digested samples were either analyzed directly

or after immunoprecipitation (IP) via the FLAG tag. Red arrowheads indicate fragments protected from PK in the absence, but not presence of detergent. The PK-protected fragment from the C-terminally tagged Asterix was recovered by IP, suggesting that the C-terminus is located within the ER lumen and the N-terminus is located in the cytosol. The relative size difference between the N-and C-terminally tagged constructs observed after PK digestion can be attributed to digestion or protection of the FLAG tag. Below the gel is a cartoon depiction of one possible topology based on the results and the protease protected fragments that remain after digestion with PK. The other possible topology is a single-spanning orientation with HD2 and HD3 in the lumen. **(C)** Schematic of human Asterix with a C-terminal FLAG tag in its predicted 3-TMD topology based on the protease digestion results in panel B. To test this prediction, a cysteine-free version of Asterix (No Cys) was modified with single cysteines at the position indicated by the red-asterisks. If the topology prediction is correct, only the N-terminal domain (NTD) cysteine and the Loop 2 cysteine should be accessible to sulfhydryl modifying reagents added to the cytosolic side of them membrane. If the protein only spans the membrane once with the N-terminus facing the cytosol, then the Loop 2 cysteine should not be modified. As shown in Fig. 3.12B, wild type Asterix naturally has four cysteines, only one of which should be exposed to the cytosol because it is in the NTD. **(D)** Asterix KO HEK-293 cells were transiently transfected with the indicated Asterix-FLAG constructs, semi-permeabilized in 0.01% digitonin, washed to remove digitonin, and treated with PEG-Maleimide (average molecular weight 5 kDa) in order to modify any cytosolically exposed cysteine residues. WT Asterix contains 4 native cysteine residues, one in the N-terminus preceding TMD1 and 3 others within the putative TMD regions. Modification was only observed for the NTD cysteine and the cysteine in Loop 2, supporting a 3-TMD topology as depicted in panel B and C. The single cysteine present in the cytosolic domain of Sec61β was used as a positive control demonstrating equal modification efficiency in all samples, and the No Cys construct verifies sulfhydryl-dependent modification. Protection of the cysteines in TMD1 and TMD2 from modification verifies that membrane integrity was maintained in the experiment. All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 25% of the total reaction.

Asterix is the substrate binding unit of the PAT complex.

Direct interactions between Rho TM1+2 and Asterix through BMH crosslinking firmly establish Asterix as having substrate binding capacity. The contribution of CCDC47 to this interaction remains unclear due to the absence of cysteines in its TMD. However, the absence of any highly conserved residues within the entire N-terminus of CCDC47, including the bulk of the TMD region (Fig 3.12), argues strongly against any direct role in substrate engagement within the membrane. To better define the native PAT complex-substrate interactions, we probed direct interactions of the β_1AR TMD1 using site-specific incorporation of the photoactivatable non-natural amino acid, benzoyl-phenylalanine (Bpa) (Fig. 3.14A) (Lin et al., 2019). Initial characterization of UV-activated crosslinking confirmed complete translation and Bpa incorporation of an amber suppressed prolactin RNC (Fig. 3.14B). Additionally, UV dependent crosslinking to previously characterized signal peptide

interacting partners at this nascent chain length were observed. Immunoprecipitations confirmed crosslinks to SRP54, TRAM, TRAPα, and Sec61α. This provides a highly specific system for further characterization of TMD-PAT complex interactions.

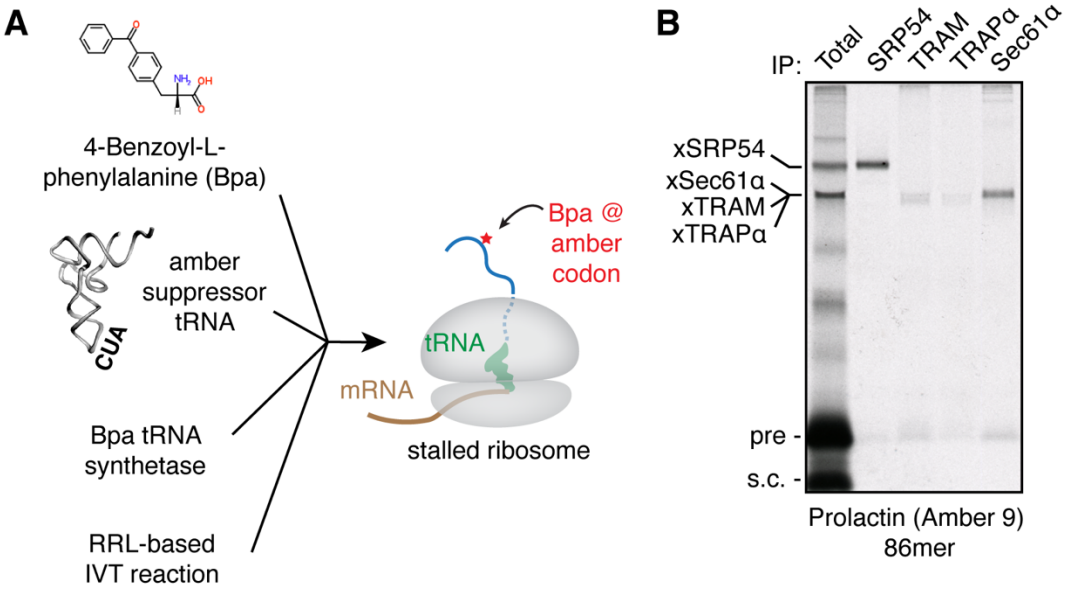


Figure 3.14 – A system to site-specifically probe PAT complex interactions. (A) Schematic of the strategy for site-specific incorporation of the photo-crosslinking amino acid Benzoyl-phenylalanine (Bpa) during in vitro translation (IVT). Bpa, a synthetic amber-suppressor tRNA, and recombinant Bpa tRNA synthetase are added to an IVT reaction. The nascent protein that is produced incorporates Bpa at an amber codon. UV irradiation results in Bpa activation and crosslinking to adjacent proteins. **(B)** The photo-crosslinking strategy was tested using a well-validated translocation intermediate: the 86mer of the secretory protein prolactin. The amber codon was installed at position 9, within the hydrophobic core of the signal sequence. At this length, the majority of the nascent chain is precursor (pre), with a small amount that is signal-cleaved (s.c.). The primary crosslinks to SRP54 and components of the translocation site (Sec61α, TRAM, and TRAPα) were verified by immunoprecipitation. All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 20% of the total reaction.

We initially monitored the interactions of β_1AR TMD1 with Bpa incorporated at a single position (aa-52), near the middle of the TMD (Fig. 3.15A). RNCs were generated by in vitro translation at a length in which the TMD is 75aa from the PTC. This allows the TMD to effectively insert and migrate into the lipid bilayer, away from the Sec61 translocon, explaining why no crosslinks to Sec61α were observed. Additionally, several prominent UV specific crosslinks were accounted for

with subsequent immunoprecipitations under denaturing conditions. Unsurprising, these were identified as SRP54 and another well-established TMD chaperone, Ubiquilin 2 (UBqln2) (Itakura et al., 2016). Notably, the intensity of these crosslinks diminished significantly when insertion efficiency increased, indicated by an increase in the glycosylated product (Fig. 3.15B, middle panel). The only other prominent interaction with TMD1 was verified to be Asterix and no appreciable interaction between the nascent TMD and CCDC47 was observed. Subsequent experiments incorporating Bpa at 11 different positions along the TMD confirmed no TMD-CCDC47 crosslinks (Fig. 3.15B) Additionally, these experiments permitted us to define the boundaries of the Asterix engagement (Fig. 3.15B-C). Consistent with the relatively short predicted Asterix TMDs (Fig. 3.15), Bpa positioned towards the edges or outside the TMD region did not form crosslinks with Asterix (Fig. 3.15B-C). Conversely, positions within the middle of the TMD readily formed crosslinks. Importantly, 4 sequential positions provided a complete 360-degree accounting of the TMD interactions with Asterix (Fig. 3.15C). Thus, consistent with the chemical crosslinking experiments, Asterix is the substrate-interacting subunit of the PAT complex, while CCDC47 is needed for Asterix stability

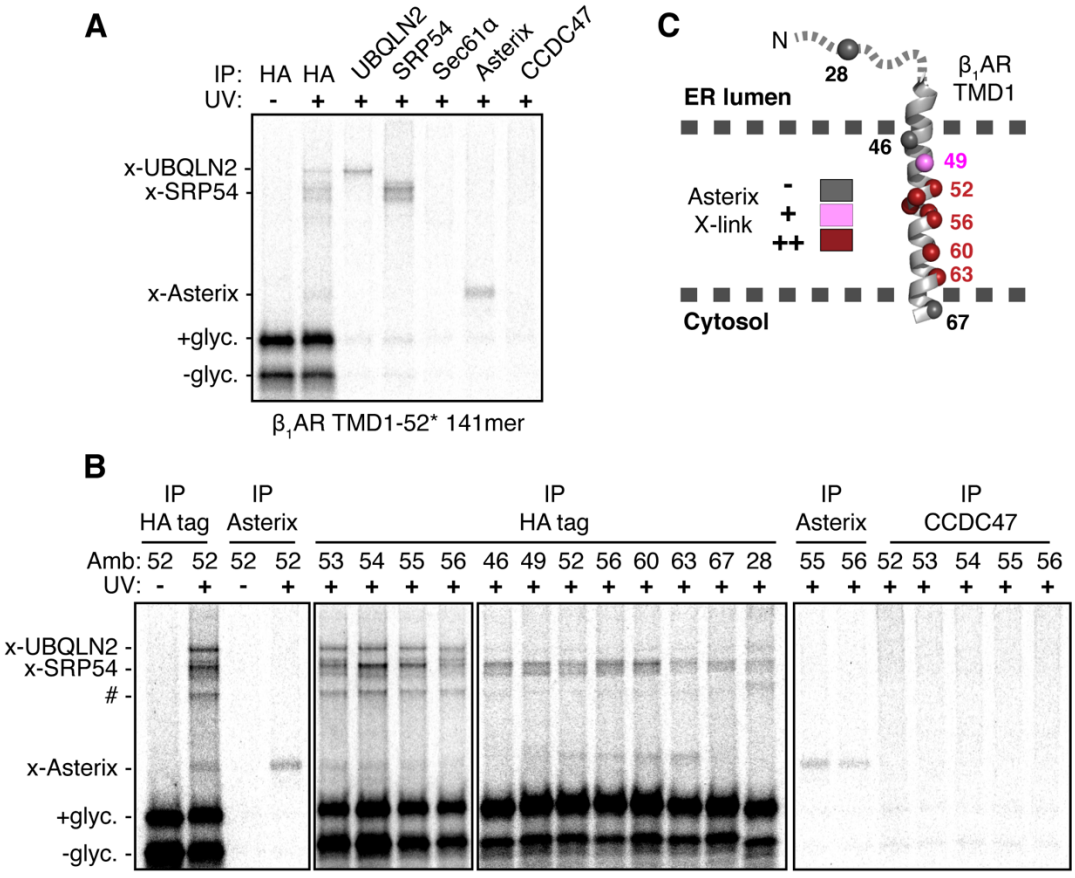


Figure 3.15 – Asterix is the substrate binding subunit of the PAT complex. (A) Site-specific photo-crosslinking of a 141mer RNC containing the UV-activated photo-crosslinking amino acid benzoyl-phenylalanine (Bpa) at position 52 within TMD1. Total translation products recovered by IP via an HA tag on the nascent chain are shown adjacent to parallel IPs using the indicated antibodies. RNCs that failed to engage SRP crosslink to UBQLN2, a quality control factor that binds exposed TMDs. A subset of RNCs fail to release from SRP and crosslink to SRP54. Of the membrane-inserted RNCs, the main crosslink is to Asterix. At this length, the TMD has moved away from Sec61α, so crosslinks to this factor are minimal. No crosslinking to CCDC47 were ever observed. **(B)** Site-specific photo-crosslinking of a 141mer RNC containing the UV-activated photo-crosslinking amino acid benzoyl-phenylalanine (Bpa) at the indicated amber positions (Amb). The diagram above the autoradiographs shows a schematic of the construct with the appropriate amino acid numbering. Amino acids in red show the strongest crosslinks to Asterix, pink show reduced crosslinks, and grey no detectable crosslinks. Total translation products recovered by IP via an HA tag on the nascent chain are shown adjacent to parallel IPs of selected samples using the indicated antibodies. Although not all IPs are shown, each position was tested for crosslinking to Asterix and CCDC47. RNCs that failed to engage SRP crosslink to UBQLN2, a quality control factor that binds exposed TMDs. Note that this crosslink diminishes markedly when increased RMs are used in the reaction (lanes 9-16, compared to lanes 5-8), presumably because the RMs contribute SRP, which is otherwise limiting in the reaction. A subset of RNCs fail to release from SRP and crosslink to SRP54. The crosslink indicated by the hashtag (#) is likely to be a mixture of similarly migrating crosslinks. The middle panel (lanes 9-16) represents a sseparate experiement in which 2x the amount of microsomes were added to the reaction mixture. As a result the insertion efficiency is higher in this experiment as can be observed by the qualitative

ration between the glyc. and unglyc. species. Because this crosslink diminishes substantially with increased RMs (see lanes 9-16) similar to the UBQLN2 crosslink), it is likely to be SGTA, another TMD-binding factor in the cytosol of this size. A small proportion of this crosslink could be the similarly sized Sec61α or TRAM. Of the membrane-inserted RNCs, the main crosslink is to Asterix, seen prominently for residues 52 to 63. At this length, the TMD has moved away from Sec61α, so crosslinks to this factor are minimal. No crosslinking to CCDC47 were ever observed. Note that by testing five sequential positions in the center of the TMD, all sides of the helix have been sampled. **(C)** Summary of Asterix crosslinks observed (or not) from different positions in or near TMD1 of β1AR. No CCDC47 crosslinks were seen from any of these positions. All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 15% of the total reaction.

The PAT complex facilitates multipass membrane protein biogenesis

Using the dual-color ratiometric assay for protein stability (see Fig. 2.1), we found that CRISPR-mediated disruption of CCDC47 or Asterix in HEK293 cells led to reduced levels of the GPCR angiotensin Type-II receptor 2 (AGTR2) relative to an internal control for translation levels (Fig. 3.16). Neither the tail-anchored membrane protein squalene synthase (SQS) nor the type II single-pass membrane protein asialyloglycoprotein receptor (ASGR) were affected in the knockout cells using the same assay. These controls indicate that the TMD insertase EMC (used by SQS (Guna et al., 2017) and the first TMD of many GPCRs (see chapter II))and the Sec61 complex (used by ASGR (Görlich and Rapoport, 1993)) are functioning correctly in cells lacking the PAT complex. This is consistent with immunoblotting revealing no substantial alterations in the expression of these and other biogenesis factors including the oligosaccharyl transferase, CNX, and Sec63 (Fig. 3.12C).

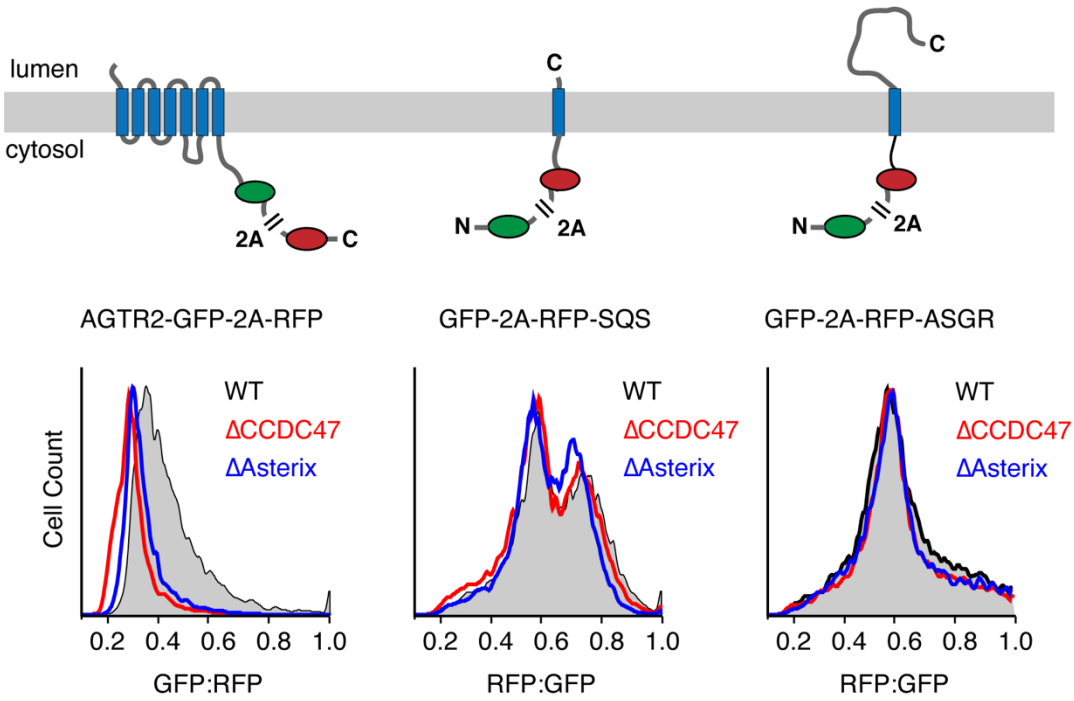


Figure 3.16 - The PAT complex is required for multipass membrane protein biogenesis. The diagrams depict dual-color fluorescent reporters for protein stability as an indirect measure of successful biogenesis. The membrane protein of interest is tagged with one fluorescent protein (FP), which is separated from a second FP by the viral 2A peptide sequence. When the 2A sequence is translated, peptide bond formation is skipped without perturbing elongation by the ribosome. Thus, translation results in two separate proteins made in a 1:1 stoichiometry that are separated at the 2A sequence. If biogenesis of the membrane protein is impaired, it will be degraded along with its tagged FP, resulting in an altered ratio of the two FPs. Thus, treatment conditions that impair biogenesis of the membrane protein will be reflected as a relative change in the ratio of FPs. The three reporters encoding angiotensin type-II recepor 2 (AGTR2), squalene synthase (SQS) and Asialglycoprotein receptor (ASGR) were transiently transfected into wild-type (WT), CCDC47 KO (ΔCCDC47) or Asterix KO (ΔAsterix) HEK293 cells and analyzed by dual-color flow cytometry. Histograms represent the distribution of FP ratio in WT (black), ΔCCDC47 (red) and ΔAsterix (blue) cells. A biogenesis defect is only seen for the multipass membrane protein AGTR2, but not for the tail-anchored protein SQS or the signal-anchored single pass protein ASGR.

Because these knockout cell lines were unstable and seemed to adapt over time, we used acute PAT complex depletion to analyze a broader range of membrane protein reporters stably expressed from a single, doxycycline inducible promoter. This strategy allowed us to acutely challenge PAT complex depleted cells with newly synthesized multipass membrane proteins. Asterix depletion, which co-depletes CCDC47 (Fig. 3.17), had minimal effects on SQS but showed reduced biogenesis of AGTR2 (as seen with knockout cells) and rhodopsin (Fig. 3.18A), which was used to characterize the PAT complex interaction. We further evaluated the broad necessity of the PAT complex using a sample population of membrane proteins with varying topologies, biophysical characteristics and number of TMDs. Three other GPCRs demonstrated an increase in post-translational degradation upon depletion of the PAT complex (Fig. 3.18B), these included the beta-1-adrenergic receptor (β_1AR), melatonin-related receptor (MTR1L), and the muscarinic acetylcholine receptor M1 (CHRM1). Importantly, reduced biogenesis was also seen for the ER resident protein TRAM2 and the cation channel ANO6, two unrelated proteins with different topology and folding demands (Fig. 3.18C). Furthermore, constraining the insertion and topology of AGTR2 by addition of an N-terminal signal peptide (SS-T4L- AGTR2) did not bypass the necessity of the PAT complex (Fig. 3.18C). This is in stark contrast to the requirement for EMC and argues strongly that while both factors are necessary for biogenesis, they act at two distinct steps. Thus, the PAT complex is needed for optimal biogenesis of various multipass membrane proteins and is likely acting at a later step (after accurate insertion occurs). This result explains why both genes were recovered in a recent screen for factors that impair surface expression of a mutant TRP6 channel (Talbot et al., 2019).

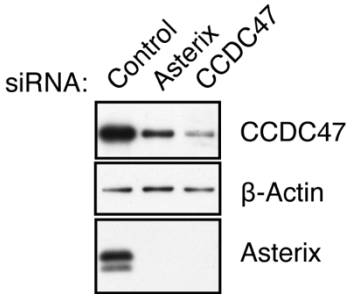
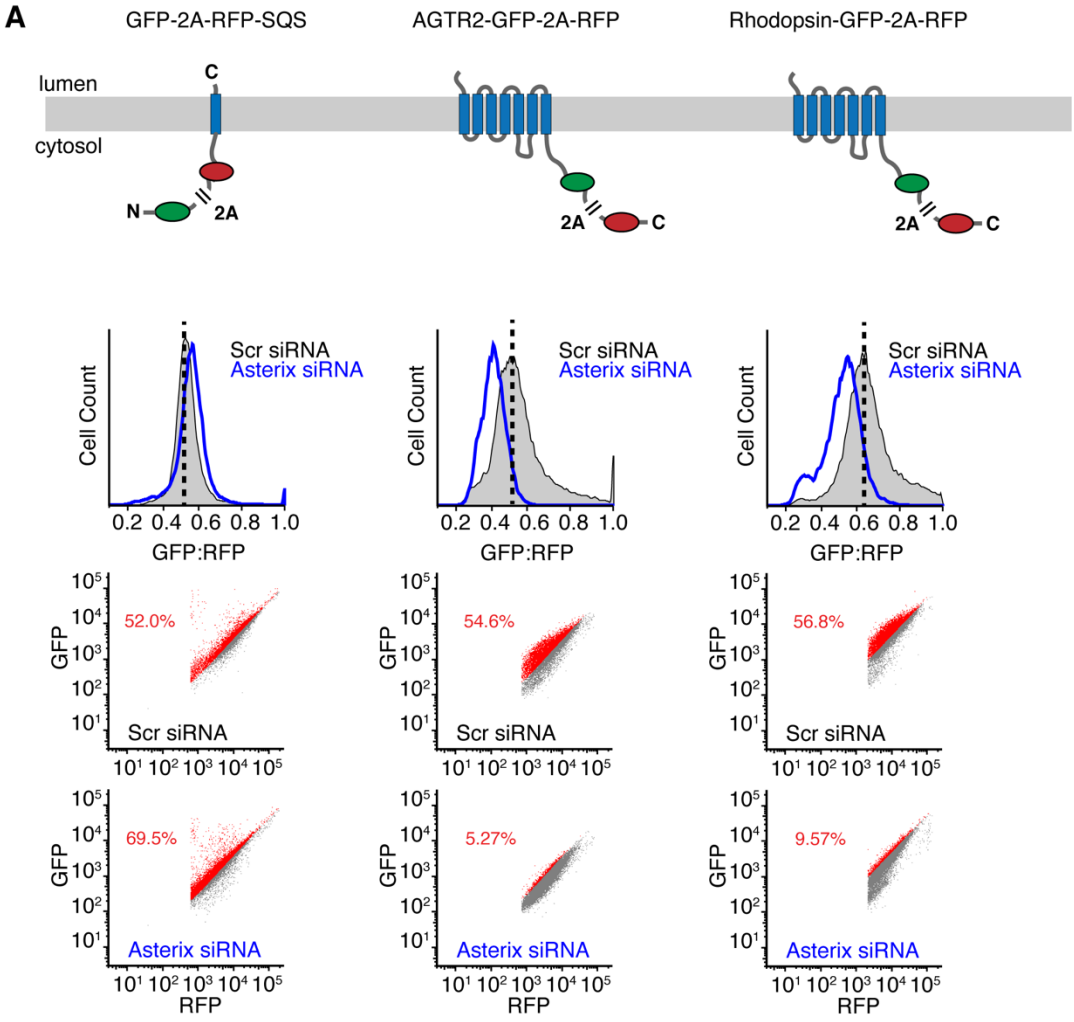
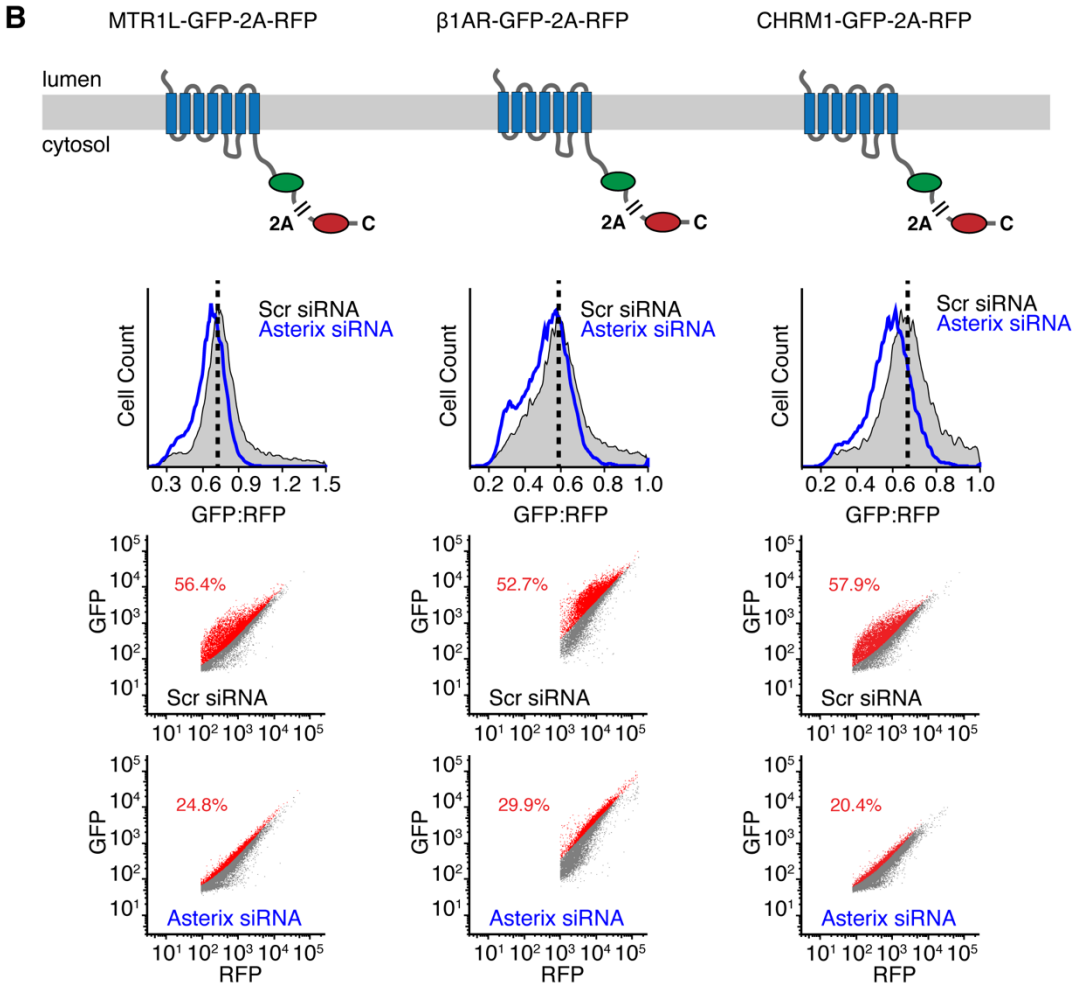


Figure 3.17 – siRNA depletion of PAT complex. Representative blot of negative control siRNA treated cells compared to cells treated with Asterix or CCDC47 siRNA.


A
GFP-2A-RFP-SQS
AGTR2-GFP-2A-RFP
Rhodopsin-GFP-2A-RFP
lumen
cytosol
C
N
2A
2A
C
2A
C
Cell Count
Scr siRNA
Asterix siRNA
0.2 0.4 0.6 0.8 1.0
GFP:RFP
GFP
RFP
52.0%
54.6%
56.8%
69.5%
5.27%
9.57%
Scr siRNA
Asterix siRNA


B
MTR1L-GFP-2A-RFP
β1AR-GFP-2A-RFP
CHRM1-GFP-2A-RFP
lumen
cytosol
2A
C
Cell Count
Scr siRNA
Asterix siRNA
GFP:RFP
GFP
RFP
56.4%
52.7%
57.9%
24.8%
29.9%
20.4%

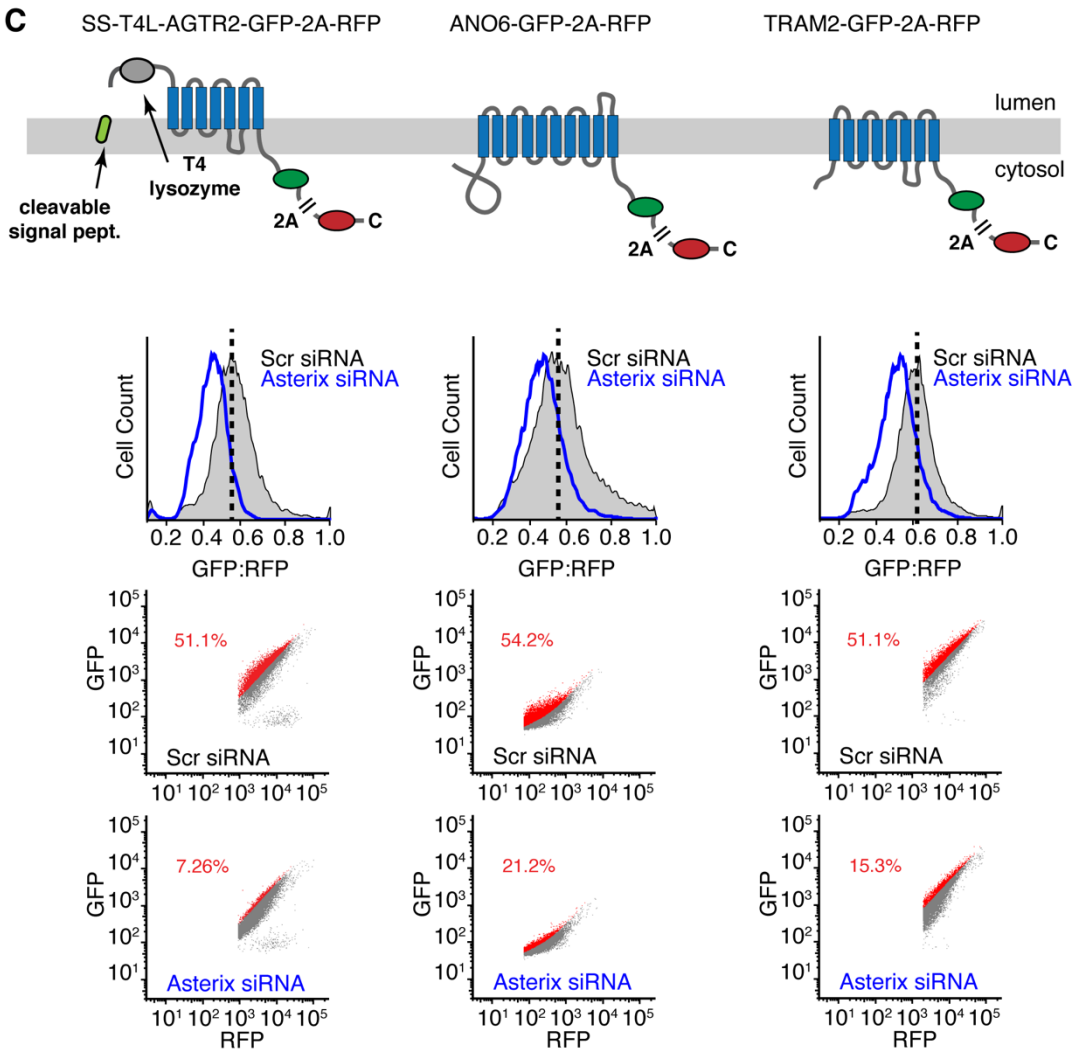


Fig. 3.18 - The PAT complex facilitates biogenesis of multipass membrane proteins. (A, B, C) Each set of proteins analyzed contains a cartoon depicting the topology, number of TMDs, and fluorescent proteins for the membrane protein reporters tested in a panel above the data panels. The viral P2A peptide sequence results in two proteins from a single translation reaction as indicated. Stable cell lines containing the indicated inducible reporters were treated with scrambled (Scr) or Asterix-targeting siRNAs, reporter expression was induced for ~6 hours, and the cells were analyzed by flow cytometry. The top plot in each column shows the histogram of fluorescent protein ratios in control cells (grey) and Asterix-knockdown cells (blue). The dashed black line indicates the mode for the control population and defines a gate used for quantification. The bottom plots show individual cells with those above the gate in red and below the gate in grey. The percent of cells above the gate is indicated.

TMDs with polar residues engage the PAT complex

Analysis of substrate insertion and crosslinking in microsomes lacking the PAT complex showed that TMD1 insertion is unimpaired (Fig. 3.10), unlike the failed insertion phenotype seen when microsomes lack the EMC insertase (see Chapter II). Furthermore, the timing of TMD1 interaction with and release from Sec61 was not appreciably changed by PAT complex depletion (Fig. 3.19). Thus, the PAT complex does not participate in targeting and is neither an insertase nor a facilitator of TMD release from Sec61α, explaining why single-pass membrane proteins like SQS and ASGR are unaffected by PAT complex depletion. Instead, the PAT complex acts after a TMD accesses the lipid bilayer.

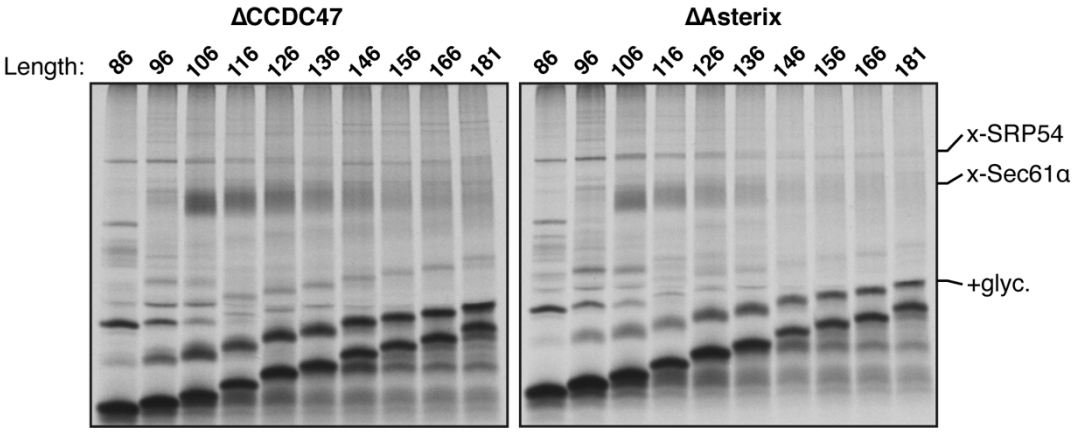


Figure. 3.19 - The PAT complex is not required for TMD1 insertion. Rho TM1+2 RNCs of varying nascent chain lengths (indicated at top of gels) were translated in vitro in the presence of RM prepared from ΔCCDC47 or ΔAsterix HEK293 cells as indicated. Membranes were isolated by centrifugation through a sucrose cushion and treated with the chemical crosslinking reagent BMH. The samples were denatured in 1% SDS and immunoprecipitated via the N-terminal FLAG tag. Notice that the glycosylation of substrate is very similar in efficiency and timing as that seen in RM prepared from wild type HEK293 cells (see Fig. 3.2 for comparison). Furthermore, the appearance and disappearance of the SRP54 and Sec61α cross linking adducts are not appreciably altered from the results seen in RM prepared from wild type cells. Thus, the early steps of Rhodopsin biogenesis are not impaired appreciably in the absence of Asterix or CCDC47. As expected, the crosslink to Asterix/PAT10 is not seen. Crosslinking products seen at the approximate size of the Asterix crosslink are therefore other protein(s).

Replacement of the sole polar amino acid (asparagine) in TMD1 with leucine markedly reduced the Asterix-TMD1 interaction at all nascent chain lengths that were examined (Fig. 3.20A-B). Reintroduction of polar or charged residues at other positions in the TMD partially restored crosslinking to Asterix (Fig. 3.20C). These data indicate Asterix engages TMDs that expose hydrophilic residues within the lipid bilayer in a mostly position-independent manner. Dependence on charged residues for PAT complex association was corroborated via native immunoprecipitation (Fig. 3.20D), ruling out altered cysteine reactivity/accessibility as the explanation for the reduced crosslinking efficiency seen in figure 3.30C. This explains previous observations that PAT10 can crosslink with TMDs of either orientation and of different unrelated sequences (Meacock et al., 2002).

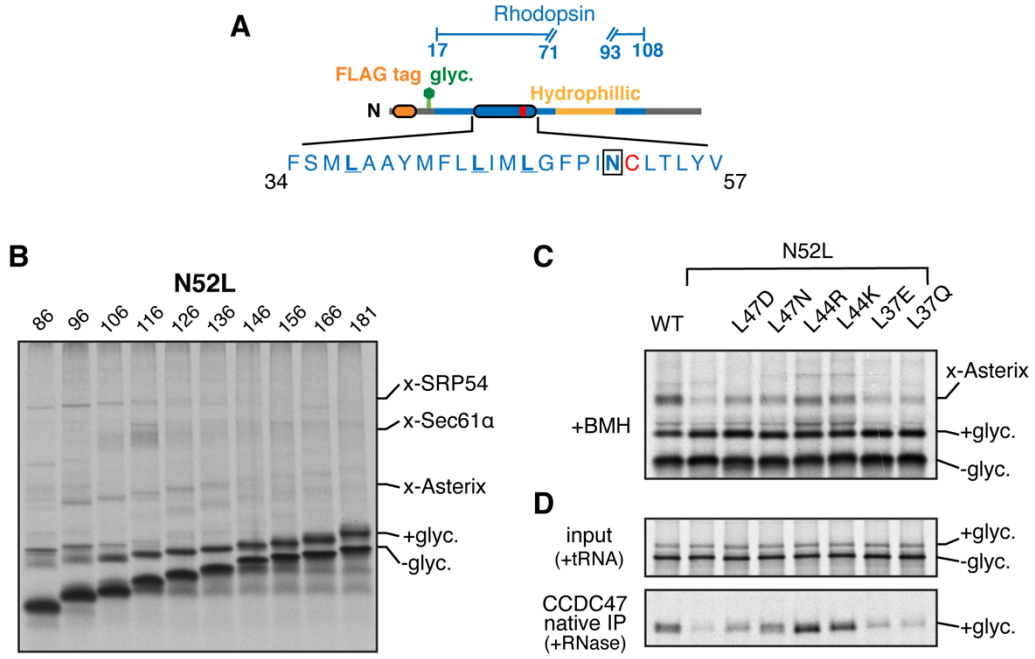


Figure 3.20 - The PAT complex engages TMDs via exposed polar residues. (A) A diagram of the Rho TM1 construct diagram containing the N52L mutation and further reintroduction of new polar residues (underlined). **(B)** Analysis by chemical crosslinking of the N52L construct depicted in A. Compared to the identical construct without the N52L mutation (Fig. 3.2), crosslinks to Asterix are markedly diminished and strong crosslinks to Sec61α are only seen for the 118mer. This is consistent with a more hydrophobic TMD leaving the translocon and migrating into the lipid bilayer faster. **(C)** Crosslinking reactions of 146mer RNCs of Rho TM1 containing the indicated mutations. **(D)** Translation reactions as in panel B (but without crosslinking) were either analyzed directly (input) or subjected to native IP using anti-CCDC47 antibodies. Nascent chains were released from the ribosome with RNase A before IP. The glycosylated substrate recovered with CCDC47 was visualized by autoradiography. Note that the efficiencies of crosslinking to Asterix in panel C corelate to the efficiencies of

recovery with CCDC47. All immunoprecipitations were loaded such that the total represents 10% of the total reaction and the IP represents 25% of the total reaction.

Substrate folding triggers PAT complex dissociation

Although the PAT complex initially engages TMDs co-translationally as a peptidyl-tRNA, the complex remains associated even after termination releases the nascent chain from the ribosome-Sec61 complex (Fig. 3.21A). This observation, together with the finding that exposed hydrophilic residues within a TMD influence PAT complex interaction, suggested PAT complex dissociation might be triggered when TMDs pack correctly to shield exposed hydrophilicity. To test this, we determined whether PAT complex interaction is selective to immature but not a folded membrane protein. Exploiting the fact that an engineered β_1AR folds efficiently after in vitro translation (see chapter II), we compared its interaction with the PAT complex relative to truncation products intended to mimic biogenesis intermediates. A pooled mixture of intermediates truncated after each TMD produced by in vitro translation was subjected to native IP using anti-CCDC47 antibodies. Full length β_1AR was recovered substantially less efficiency than each of the intermediates, suggesting the PAT complex dissociates upon correct folding of β_1AR (Fig. 3.21B). Furthermore, this result was corroborated when probing the interaction of an intermediate product vs. a full-length β_1AR, in native membranes, using photo-crosslinking of a BpA placed in the first TMD (Fig. 3.21C). In vitro translation of full-length β_1AR inherently produces intermediate products which either represent stalled translation, or premature termination (Figure 3.21C, see blue stars). Satisfyingly, these intermediate products appear to crosslink to Asterix and can be recovered via immunoprecipitations under denaturing conditions. Conversely, crosslinks to the full length β_1AR are not recovered despite similar levels of total translated product (Fig. 3.21C, see red stars). Thus, the PAT complex engages partially translated products to shield hydrophilicity from the surround lipid environment and releases substrate upon complete folding. Considering all previous observations, this defines the PAT complex as an intramembrane molecular chaperone.

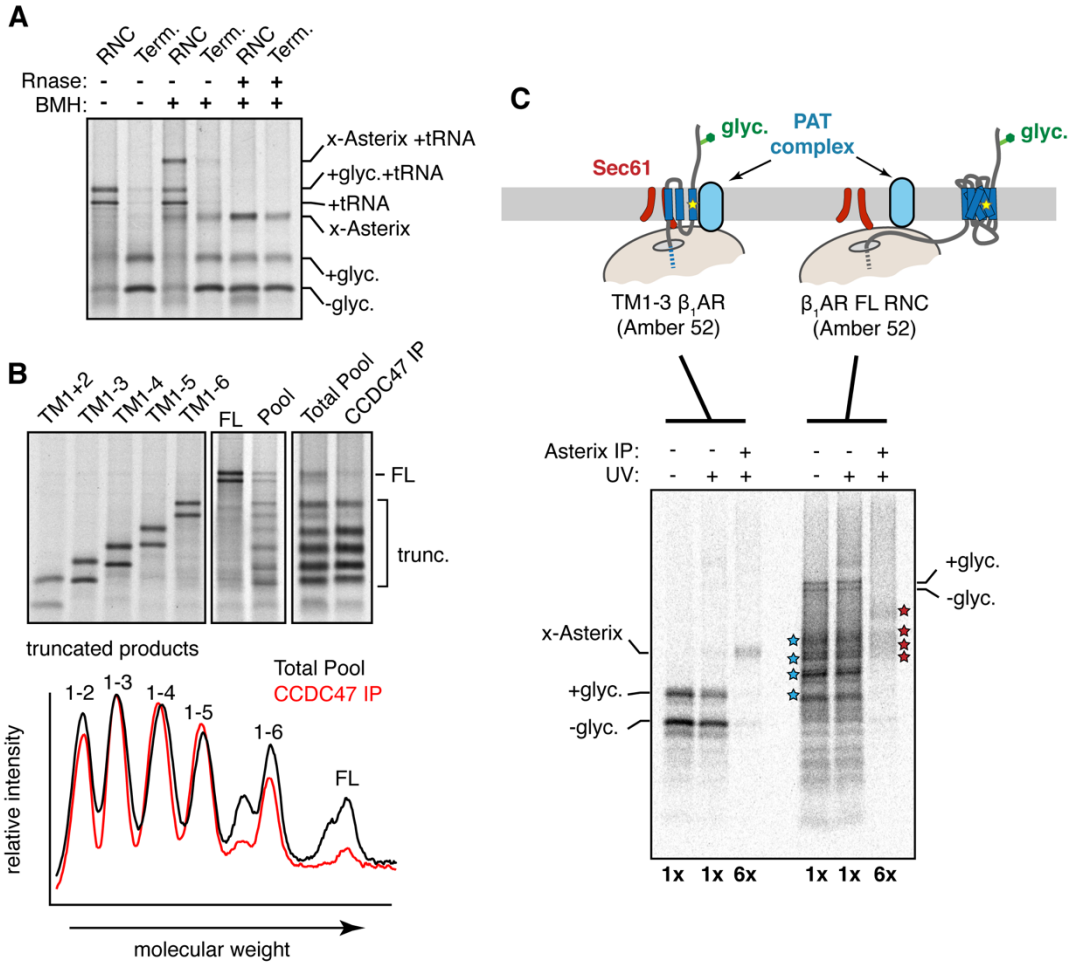


Figure 3.21 The PAT complex releases substrate upon folding. (A) Terminated (Term.) or truncated (RNC) Rho TM1+2 was inserted into RMs and treated with BMH where indicated. Before SDS-PAGE, some of the samples were digested with RNase A as indicated to remove the tRNA. **(B)** Full length β1AR (FL) or constructs truncated after each TMD were translated in the presence of RMs and analyzed either individually (left panel) or pooled before analysis. All constructs contained a stop codon and are terminated. The membrane-targeted population from the pooled reaction was isolated by sedimentation and divided in two aliquots. One aliquot was set aside (Total Pool) and the other was used for a native IP with anti-CCDC47 antibody. Both samples were then subjected to denaturing pulldown via the C-terminal His tag to ensure that only completed translation products were visualized. The graph below the gel represents scanning densitometry of the last two lanes. Note that substantially less full length β1AR is recovered with CCDC47 relative to each of the truncation products. **(C)** Top cartoon depicts the experimental strategy for comparing Asterix interaction with a membrane protein intermediate versus full length product. In this experiment, the photo-crosslinking amino acid Bpa (yellow star) is incorporated into position 52 within TMD1 of β1AR by in vitro translation. The intermediate is represented by the TM1-3 product containing the first three TMDs of β1AR. The full length (FL) β1AR contains all seven TMDs followed by a long flexible linker. TM1-3 is stalled 35 amino acids downstream of TMD3 (with TMD4 inside the ribosomal tunnel), allowing TMD3 to be outside the ribosome. β1AR FL is stalled 152 amino acids downstream of TMD7, providing a sufficiently long tether for all seven TMDs to have emerged, inserted into the membrane, and assembled together. The translation products are then irradiated with UV light to activate the Bpa and any crosslinking to Asterix is subsequently detected by denaturing IP via Asterix. The microsomes from

the IVT reaction were isolated, resuspended, irradiated with UV light (or left untreated), and denatured. The samples were then divided in two and immunoprecipitated via the nascent chain or via Asterix. Six-fold more of the Asterix IPs were loaded on the gel relative to the nascent chain IPs. As expected, the Bpa in TMD1 crosslinks to Asterix in the TM1-3 intermediate. Note that the crosslinked band in the 6x Asterix IP is the same intensity as the glycosylated band in the 1x nascent chain IP. Although elongation to the full-length product was somewhat inefficient, clear glycosylated and non-glycosylated products are observed in the nascent chain IPs. No band is seen in the 6x Asterix IP sample that is of comparable intensity to the glycosylated band in the 1x nascent chain IP. This argues that the proximity of TMD1 to Asterix has diminished substantially in the full-length nascent chain relative to the TM1-3 intermediate. Of note, a heterogeneous set of crosslinks (marked by red stars) are seen at a lower part of the gel in the 6x Asterix IP. These correspond to the sizes expected for Asterix crosslinks (i.e., shifted by ~10 kD) to the major incomplete translation products (marked by blue stars). These crosslinks provide an internal control and further supports the conclusion that incomplete products engage Asterix, while a complete 7-TMD product does not.

A model for the PAT complex as an intramembrane chaperone

Our findings indicate that the PAT complex is a widely conserved ER resident complex that directly interacts with nascent TMDs, releases its substrates upon correct folding, and is needed for optimal membrane protein biogenesis. These features define the PAT complex as an intramembrane chaperone. Its preference for TMDs with exposed hydrophilic amino acids within the lipid bilayer is analogous to soluble chaperones that prefer hydrophobic patches exposed to the aqueous environment (Blond-Elguindi et al., 1993; Döring et al., 2017; Stein et al., 2019). The more favorable TMD-TMD interactions that accompany correct folding likely displaces the PAT complex, whose interaction must necessarily be more generic and hence weaker. Just as the binding cleft of chaperones like Hsp70 are hydrophobic (Mayer and Bukau, 2005), the substrate-interaction domain of Asterix may be hydrophilic, consistent with a large number of conserved hydroxylated amino acids in its TMDs.

Analogous intramembrane interactions with hydrophilic TMDs have been proposed to facilitate recognition of misfolded membrane proteins during quality control (Natarajan et al., 2019; Sato et al., 2009, 2003). By binding to and shielding similar motifs in a nascent protein, the PAT complex probably prevents promiscuous degradation. This would explain why its depletion leads to post-translational reduction of various multipass membrane proteins. The PAT complex may have priority over quality control factors for two reasons. First, the PAT complex is highly abundant, comparable to or even exceeding the Sec61 complex (Ghaemmaghami et al.,

2003; Itzhak et al., 2016). Second, the cytosolic domain of CCDC47, which is enriched in basic amino acids, may have some affinity for the ribosome to recruit the PAT complex in vicinity of the Sec61 complex. This interaction is likely to be weak and dynamic as the PAT complex has not been seen in proteomic analysis of ribosome translocon complexes (Conti et al., 2015).

Even though there are severe phenotypes associated with loss of CCDC47 at the level of an organism and in humans (Morimoto et al., 2018; Yamamoto et al., 2014), loss of the PAT complex is tolerated at the cellular level and substrates still mature at a reasonable efficiency. Many proteins can fold without assistance from chaperones, whose primary role is to reduce off-pathway interactions and maximize productive folding (Ellis and van der Vies, 1991). Furthermore, individual chaperone disruptions often have only modest cellular phenotypes due to their considerable redundancy (Wang et al., 2015). Thus, there may be many other intramembrane chaperones that can partially compensate for the PAT complex in its absence. While several substrate-specific chaperones have been proposed for particular membrane proteins (Brechet et al., 2017; Gu et al., 2016; Kota and Ljungdahl, 2005), our findings suggest that the PAT complex is an abundant generalist for semi-hydrophilic TMDs that are found in nearly all of the over 2500 multi-pass membrane proteins in humans.

Chapter III: An intramembrane chaperone facilitates multipass membrane protein biogenesis

Chapter IV

A model for multipass membrane protein biogenesis

The cooperative action of insertion machineries and an intramembrane chaperone.

In this thesis I have provided two significant insights into the mechanisms governing multipass membrane protein biogenesis. The first involves the long-standing question in the field of how topogenic determinants are recognized and decoded by cellular machinery to accurately insert membrane proteins in the correct orientation. The second insight illuminates the action of a new intramembrane chaperone complex (the PAT complex) acting to offset destabilizing features of nascent TMDs until complete synthesis and folding.

Our studies demonstrate N_{exo} signal anchors use EMC for insertion, while N_{cyt} TMDs and signal sequences go through the Sec61 complex. Thus, the mechanism involves two separate machineries for the insertion of the first hydrophobic element, determining topology. This dichotomy is supported by our observations that a range of N_{exo} TMDs do not require Sec61 for their insertion using in-vitro insertion assays (Figs. 2.19, 2.20. 2.21) but demonstrate a strong dependence on EMC both in-vitro and in-vivo (Figs. 2.14, 2.15, 2.16). Consistent with previous mutagenesis assays (von Heijne, 2006), N_{exo} TMDs that are short, hydrophobic and contain cytosolically localized positive charges can be tolerated by Sec61. This explains why the classical N_{exo} substrate, leader peptidase (LepB), could be inserted in a purified system (Görlich and Rapoport, 1993). However, extension of the LepB TMD by three additional amino acids rendered it completely EMC dependent (Fig. 2.17). Furthermore, this model nicely reconciles previous observations in which potent Sec61 inhibitors, that block opening of the lateral gate, only inhibit insertion of N_{cyt} TMDs and signal sequences, but not N_{exo} signal anchors (Luesch and Paavilainen, 2020).

Fundamentally, the major remaining question regarding EMC function is when exactly it acts on substrate during co-translational insertion. Considering our recent findings suggesting two separate paths for insertion (described above), we can make a reasonable inference that the observed topological inversion of β_1AR-TMD1(in the absence of EMC) is a result of Sec61-mediated insertion. This necessarily means EMC must be acting at a step prior to ribosome docking on the translocon, when a TMD will be heavily biased towards engaging Sec61, and EMC action will be sterically blocked by the tight interface (Voorhees et al., 2014). Therefore, we propose

a model in which a hydrophobic element emerges from the ribosome exit tunnel and is engaged by the substrate binding domain of SRP54 (Fig. 2.10). SRP will target the RNC to the ER membrane via interaction with SRP receptor with cooperative GTP hydrolysis triggering release. At this crucial intermediate the ribosome has not engaged Sec61 and the targeting element is accessible to EMC, which is highly abundant and can readily diffuse through the ER membrane. If EMC engages a N_{exo} signal anchor with appropriate sequence features (i.e. positive cytosolic charges), the TMD will be inserted via EMC in the correct orientation. Conversely, insertion of signal sequences and N_{cyt} TMDs will be rejected and defer to Sec61. This model remains largely speculative as the mechanistic underpinnings that allow substrate selection/rejection by EMC will require structural analysis. Nevertheless, in a purified system EMC appears to only be capable of N_{exo} insertion (Fig. 2.22, 2.23) and Sec61 optimized for N_{cyt} insertion (with the exception of ideal N_{exo} TMDs).

While N_{exo} signal anchors would not obligately use Sec61 for insertion, they can probably engage Sec61 at its lateral gate after insertion. This idea is favored by Sec61's close proximity to the nascent chain via ribosome binding (Kalies et al., 1994), the signal-binding capacity of Sec61's lateral gate (Gogala et al., 2014; Li et al., 2016; Plath et al., 1998; Voorhees and Hegde, 2016), and the observed Sec61-TMD crosslinking (High et al., 1993a). The signal anchor would then be positioned ideally for interacting with the next TMD, whose insertion would occur via Sec61 (Fig. 2.24).

At this point, the first TMD can leave the immediate environment of the translocon and engage the PAT complex, if it contains destabilizing hydrophilic residues (Fig. 3.20). A single TMD would very likely be stable within the membrane as an α-helix (Krishnamani et al., 2012; Lau and Bowie, 1997), consistent with careful biophysical measurements demonstrating hundreds of ATP molecules are required to remove a TMD from the membrane (Yang et al., 2018). The primary challenge faced by a multipass membrane protein is off-pathway interactions disrupting assembly, or premature degradation by quality control mechanisms (Chen and Gouaux, 1999). Consistent with this assessment, in a completely purified system two halves of rhodopsin can be synthesized separately, but assemble into a functional whole upon reacquaintance (Huang et al., 1981). Thus, the PAT complex acts

to shield “naked” TMDs until assembly with their distant counterparts and the interaction is severed upon productive folding (Fig. 3.21).

The introduction of two new molecular players in the process of multipass membrane biogenesis greatly increases our understanding of this essential and complex process. The studies presented here will pave the way for future mechanistic dissection and a structural understanding of both EMC and the PAT complex.

Cell lines

All cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal calf serum (FCS). In cases where the cells contained a stably expressed doxycycline-inducible reporter, tetracycline-free FCS was used as well as 15 μg/ml blasticidin and 100 μg/ml hygromycin. All cell lines used in chapter II have been described and characterized previously (Guna et al., 2017). They include the following: Flp-In 293 T-Rex cells (wild type and ΔEMC6), and U2OS Flp-In cells (WT, ΔEMC5, ΔEMC5+EMC5 rescue). Stable cells lines in chapter III were generated using the FRT Flp-In system according to manufacturer's protocol (Invitrogen). Stably expressed cell lines included GFP-2A-RFP-SQS, β_1AR-GFP-2A-RFP, AGTR2-GFP-2A-RFP, SS-T4L-AGTR2-GFP-2A-RFP, TRAM2-GFP-2A-RFP, Ano6-GFP-2A-RFP, CHRM1-GFP-2A-RFP, MTR1L-GFP-2A-RFP, and Rhodosin-GFP-2A-RFP. CCDC47 and Asterix knockout cells were generated by transient transfection with the px459 plasmids containing the appropriate guide RNAs. After 24 h, transfected cells were selected for 48hrs with 1 μg/ml puromycin then trypsinized and diluted into 96-well plates at a concentration of 0.5cells/well to select for single cell colonies. After ~2weeks, single colonies were expanded and clones displaying undetectable CCDC47 or Asterix were chosen for further analysis.

Antibodies

CCDC47 antibody #1 (A305-100A) and CCDC47 antibody #2 (A305-101A) were obtained from Bethyl-laboratories. FLAG immunoprecipitations were performed using FLAG-M2 affinity gel (Sigma). Anti-HA antibody was generated in house (Itakura et al., 2016). Signal Peptide Peptidase (SPP) antibody was purchased from Bethyl Laboratories (A304-404A). Anti-Asterix (WDR83OS) antibody was purchased from Sigma/Human Protein Atlas (HPA065685). Anti-UBQLN2 antibody was clone 5F5 obtained from Sigma (WH0029978M3). Anti-SRP54 was from BD Biosciences (610940). Custom antibodies against Sec61α, Sec61β, TRAPα, and TRAM have been described and characterized previously (Fons et al., 2003). All antibodies for western blotting in figures 2.7 and 3.12 have been described previously (Chitwood et al., 2018).

Constructs

The parent β_1AR construct for in vitro translation was created by inserting the coding region of residues 20-424 of turkey β_1AR-B6m23 (Warne et al., 2009) into an SP64 based vector containing an HA affinity tag at the N-terminus and the unstructured cytosolic domain of Sec61β (residues 2-69, with the single Cysteine and predicted Glycosylation acceptor sequence mutated to Serine and Glutamine, respectively) followed by a 6-Histidine tag at the C-terminus. A glycosylation acceptor site (NGT) was introduced at residues 22-24 within the β_1AR sequence. From this parent construct, versions lacking the HA tag, glycosylation site, cytosolic loop 3 (CL3 residues 233-262), and TMD3 (residues 109-148) were generated by standard subcloning methods. β_1AR-TMD1 was created by deleting everything downstream of the beginning of TMD2 from the parent cassette. Similar approaches were used to create constructs β_1AR-TMD1-2 through β_1AR-TMD1-6. All GPCR-TMD1, B1AR TMD1 mutants, and LEP TMD mutant constructs were made by replacing the β_1AR TMD1 with the respective first TMDs of indicated GPCRs, or mutants of either β_1AR TMD1 or LEP TMD1, including up to 15aa of the N-terminal native sequence (or the entire native N-terminus where indicated) and the entire cytosolic loop 2 (CL2) sequence preceding TMD2. Any native cysteines were mutated to serine. Methionines were added where necessary by mutating hydrophobic residues to allow for efficient detection by autoradiography. The coding sequences for human ASGR1 (Görlich and Rapoport, 1993) and bacterial leader peptidase (LEP) with the TMD2 removed (Heinrich et al., 2000), were placed into an SP64 based vector containing a 3F4 epitope at the C-Terminus or 3xHA and glycosylation tag at the N-terminus, respectively. For the construction of the in-vivo β_1AR fluorescent reporter, the sequence encoding HA-β_1AR was sub-cloned into a pcDNA3.1 based vector containing a C-Terminal GFP-P2A-RFP reporter. For all other fluorescent reporters, a parent cassette was first created by sub-cloning the GFP-P2A-RFP fluorescent reporter into a pcDNA5/FRT/TO vector backbone. The coding sequences of CNR2 (NP_001832.1), AGTR2 (NP_000677.2), and ADA1A (NP_000671.2) were then inserted into this parent cassette with the GFP-P2A-RFP reporter at the C-Terminus. The coding regions for both ASGR and SQS were inserted at the 3' end of the GFP-P2A-RFP reporter within the pcDNA5 cassette. A gene block (IDT) encoding the

signal sequence of prolactin followed by an HA-epitope tag and the sequence for full-length Phage T4 Lysozyme was appended to all GPCR-GFP-P2A-RFP cassettes using Gibson Assembly (NEB). The T4 Lysozyme sequence (residues 2-161) had all native cysteines and predicted glycosylation acceptor sites mutated to serine or glutamine, respectively). Additionally, the N-terminal HA-epitope tag preceding the GPCR sequence was removed and replaced by the appended SS-HA-T4L sequence. SS-T4L-β_1AR for in-vitro expression in an SP64 based cassette was cloned in a similar manner. A gene block (IDT) encoding an HA tag and the TMD of MAN1A1 (NP_005898.2), including native N and C-terminal flanking residues (aa 33-75), was appended to the N-terminus of β_1AR in both the SP64 cassette and the pcDNA5 GFP-P2A-RFP cassette using Gibson Assembly. As indicated in the Key Resources Table, several β_1AR TM1 constructs and LEP TM1 constructs were ordered as gBlocks containing the SP6 promoter and coding sequence of interest. PCR for subsequent in-vitro transcription was carried out directly from these gBlocks. TRAM2 was PCR amplified from a human cDNA library and then inserted into the SP64 cassette using restriction cloning. Subsequently, the coding sequence of the TRAM2 mRNA was PCR amplified and inserted a parent pcDNA5-GFP-P2A-RFP cassette by Gibson Assembly. Sequences encoding 1xFLAG-Rho TM1+2 and 1xFLAG-Rho TM1+2 (F53C) were ordered as gene blocks (IDT) and inserted into a parent vector containing an SP6 promoter and the flexible N-terminus of Sec61β with a 6His tag appended to the C-terminus. A glycosylation acceptor site (NGT) was introduced at amino acids 13-16. Variants of these parent constructs with a TwinStrep tag, HA tag, or lacking a glycosylation site were generated by standard subcloning methods. Constructs encoding Rho TM1, where TM2 is replaced by a flexible hydrophilic linker (SGSGSGSGSGGSSGGMGGSGS), were ordered as gene blocks containing a 5' SP6 promoter and transcribed directly for in vitro translation. Similar methods were used for all Rho TM1+linker polar residue point mutants. Truncated versions and variants containing amber codons at specific sites in β_1AR were created by PCR-based cloning methods and site-directed mutagenesis, respectively, and verified by sequencing. Constructs encoding the mutant E.coli tyrosyl-tRNA synthetase in the pET21 vector and B. stearothermophilus suppressor $tRNA^{Tyr}$ sequence in the pRSET-A vector have been described (Lin et al., 2019). Templates for in vitro

translation of tagged Asterix constructs were ordered as gene blocks containing a 5' SP6 promoter and used directly for transcription and translation. Mammalian expression constructs for C-terminally FLAG-tagged human Asterix and the various cysteine variants were produced as gene blocks, sub-cloned into pcDNA3.1-based vectors, and verified by sequencing. For the creation of stable cells lines expressing various membrane protein reporters, the coding sequences for CHRM1 (NP_000729.2), bovine rhodopsin (NP_001014890.1), Anoctamin-6 (NP_001020527.2), and MTR1L (NP_004215.2) were PCR amplified and placed into a parent pcDNA5/FRT/TO vector backbone with a C-terminal GFP-P2A-RFP using Gibson Assembly (NEB). CRISPR-Cas9 mediated gene disruptions of CCDC47 and Asterix were performed using the pSpCas9(BB)-2A-Puro (PX459) plasmid (Addgene) containing the guide RNAs 5'-GTATGGACTGCCG-GACTCTT-3' (CCDC47) and 5'-AAGGCCGGGTTACATTCGCT-3' (Asterix).

Flow cytometry analysis

Analysis of reporter expression by flow cytometry was similar to previously described methods (Guna et al., 2017; Itakura et al., 2016). Transient transfection of fluorescent reporter constructs was performed using either Mirus TransIT 293 (for HEK293 T-Rex cells) or Mirus TransIT 2020 (for U2OS cells) according to manufacturer's instructions. In all experiments, 1µg/ml of total plasmid was transfected into a dish containing complete medium. The amount of the fluorescent reporter plasmid was titrated individually for each protein of interest based on transfection efficiency and expression levels, and a non-expressing plasmid was used to maintain equal amounts of total plasmid transfected (1µg/ml). For the U2OS ΔEMC5 rescue cells, re-expression of EMC5 was induced for 24-30 hours with 1 µg/ml of doxycycline prior to reporter plasmid transfection. For experiments using knockout cell lines, ΔCCDC47 and ΔAsterix cells were transiently transfected with 1 µg/ml of the appropriate pcDNA5 expression constructs ~24 h before induction with 100 ng/ml doxycycline. All transfections were performed using Mirus-Transit 293 according to manufacturer's instructions. For experiments using stably expressed reporter cell lines, siRNA depletion was performed over a ~96 h period using the Lipofectamine RNAiMAX reagent according to manufactures instructions (Thermo). Briefly, a first

round of siRNA treatment was performed in the presence of DMEM and 10% tetracycline-free FCS. Cells were incubated for 48 h before a second round of siRNA treatment was performed under the exact same conditions. Following a second ~48 h incubation, expression of fluorescent reporter constructs was induced with 100 ng/ml doxycycline for 6 h before analysis by flow cytometry. Acute expression of reporters was essential to accurately monitor degradation of reporter constructs and avoid saturation of degradation pathways. In all experiments the cells were collected by trypsinization, washed once in ice-cold PBS, then resuspended in the equivalent culture volume of PBS and 1µg/ml DAPI stain (Thermo). Cells were passed through a 70µm filter before flow cytometry analysis using a Beckton Dickinson LSRII instrument. 20,000 GFP positive (or RFP positive for SQS and ASGR) were collected. Further gating for live cells (negative for DAPI stain) and relatively high soluble fluorescent protein was used to focus on the population of cells with productive translation of reporter constructs.

In vitro transcription and translation

In vitro transcription was performed with SP6 polymerase as described and PCR products were utilized as the template (Sharma et al., 2010). The transcription reactions were conducted with 5-20 ng/µl PCR product in 40 mM HEPES pH 7.4, 6 mM $MgCl_2$, 20 mM spermidine, 10 mM reduced glutathione, 0.5 mM ATP, 0.5 mM UTP, 0.5 mM CTP, 0.1 mM GTP, 0.5 mM CAP, 0.4-0.8 U/µl RNasin and 0.4 U/µl SP6 polymerase at 37°C. In vitro translation in RRL was as described previously in detail (Feng and Shao, 2018; Sharma et al., 2010). Translations were for 20-45 minutes at 32°C unless indicated otherwise in the individual figure legends. Translation reactions typically contained 33% by volume nuclease-treated RRL, 0.5 µCi/µl ^{35}S-methionine, 20 mM HEPES, 10 mM KOH, 40 µg/ml creatine kinase, 20 µg/ml pig liver tRNA, 12 mM creatine phosphate, 1 mM ATP, 1 mM GTP, 50 mM KOAc, 2 mM $MgCl_2$, 1 mM reduced glutathione, 0.3 mM spermidine and 40 µM of each amino acid except methionine. The transcription reaction was added to 5% by volume to the translation reaction without further purification. For translation reactions in the presence of human cell-derived rough microsomes (hRMs), 0.25-1 µl of hRMs (at concentration that gives an absorbance at 280 nm of 75) were added to a

10µl translation reaction. Each batch of hRMs was titrated in preliminary experiments to achieve equal translation levels, allowing for functional comparisons between various microsomes.

Preparation of rough microsomes

Canine pancreas-derived rough microsomes (cRM) were prepared as described previously (Walter and Blobel, 1983). Preparation of microsomes from HEK293-based cells was slightly modified from earlier protocols (Zhang et al., 2013). Briefly, ten 15 cm plates of Flp-In 293 T-Rex cells (WT or ΔEMC6) were grown to 80-100% confluency, collected in ice-cold PBS, sedimented at 500 x g for 5 min at 4°C, and washed twice in ice-cold PBS. The cell pellet was resuspended in 3 pellet volumes of ice-cold sucrose buffer (10mM HEPES, pH 7.4, 250 mM sucrose, 2 mM $MgCl_2$). Cells were lysed in the cold (4°C) by ~25-30 passes through a 26 guage needle using a 1 mL syringe. The lysates were clarified of nuclei and debris by centrifugation twice at 3,800 x g for 30 min at 4°C in a tabletop micro-centrifuge. The supernatant was centrifuged at 75,000 x g for 1 hr at 4°C in an MLA-80 rotor (Beckman Coulter). The supernatant was discarded and the resulting membrane pellet was resuspended in microsome buffer (10 mM HEPES, pH 7.4, 250 mM sucrose, 1 mM $MgCl_2$, 0.5 mM DTT). Total microsome resuspension volume was adjusted such that the absorbance at 280nm was 75. Wild type RM were also obtained from HEK293 cells grown in suspension (Expi293F cells) and prepared similarly to RM from adherent cells with a few minor modifications to adjust for the larger scale. In brief, ~2 L of cells were grown to a concentration of $5x10^6$ cells/mL then collected by centrifugation. Cell pellets were washed twice with ice-cold PBS and pooled as necessary. A ~30 mL pellet was resuspended in 60 mL of sucrose buffer (50 mM HEPES, pH 7.4, 50 mM KOAc, 6 mM $Mg(OAc)_2$, 1 mM EDTA, 250 mM sucrose, 1 mM DTT) and lysis was carried out in a glass dounce homogenizer. Lysate was cleared twice by centrifugation at 3,500 x g for 30 min at 4°C. Supernatant was recovered and underlaid with one-third the volume of sucrose cushion (1.3 M sucrose, 50 mM HEPES, pH 7.4, 50 mM KOAc, 6 mM $Mg(OAc)_2$, 1mM EDTA, 1 mM DTT) and centrifuged for 1 h at 371,000 x g (60,000 rpm) and 4°C for 1 h in the Type 70Ti rotor (Beckman). The supernatant was removed by aspiration, and the pellets were resuspended

and pooled by manual homogenization in a dounce using 6 mL resuspension buffer (250 mM sucrose, 50 mM HEPES, pH 7.4, 1 mM DTT). The final preparation was adjusted to an absorbance of 75 when measured at 280 nm in 1% SDS. Canine pancreas-derived rough microsomes (cRM) were prepared as described previously (Walter and Blobel, 1983), and were used in very few experiments where explicitly stated in the figure legends. All microsome preparations were flash frozen in liquid nitrogen and stored at -80°C.

Protease protection assays

Immediately following the translation reaction, the samples were placed on ice and 10% of the reactions were set aside for analysis by SDS-PAGE and autoradiography of total products. The remainder was subjected to protease protection digests by the addition of proteinase K (PK), added to a final concentration of 0.5 mg/ml and incubated at on ice for 50 min. To stop the digestion reaction, PMSF was added to 5 mM, incubated on ice for 2-5 min, and the entire reaction transferred to 10 volumes of boiling 1% SDS, 100mM Tris-Cl, pH 8.0. For subsequent immunoprecipitations and pulldowns, samples were diluted 10-fold in ice-cold IP buffer (1x PBS supplemented with an additional 250 mM NaCl, 0.5% TX-100, 10 mM imidazole). Subsequently, samples were added to 10µl (packed) of either Nickel-NTA resin (to capture 6His-tagged proteins), or Protein A agarose plus the appropriate antibody typically used at 1:300 dilution. Immunoprecipitations were incubated for 2 hours rotating at 4°C. Following binding, the resin was washed twice with 50-100 resin volumes of IP buffer, eluted with sample buffer, and analyzed directly by SDS-PAGE and autoradiography.

Carbonate extraction

Translation reactions were chilled on ice, layered on a sucrose cushion [20% w/v sucrose in physiological salt buffer (PSB), 100 mM KOAc, 50 mM HEPES pH 7.4, 2 mM $Mg(OAc)_2$], and centrifuged at 186,000 x g for 20 min. The membrane pellet was resuspended in 20 µl PSB, 10% was set aside as the total membrane fraction, and the remainder was diluted 100-fold in 100 mM $NaCO_3$ pH 11.5 and incubated on ice for 25min. The resulting $NaCO_3$ extracted membranes were isolated through

centrifugation in the TLA120.2 rotor (Beckman Coulter) at 70,000 rpm at 4°C for 30min. The $NaCO_3$ extracted pellet was resuspended in SDS-PAGE sample buffer. After SDS-PAGE, the gels were either exposed to detect translation products by autoradiography, or subjected to immunoblotting to assess the separation of endogenous membrane and lumenal proteins (α-Calnexin 1:5,000 or α-PDI 1:1,000).

Analysis of ribosome-nascent chain complexes

For generating templates of truncated mRNAs, PCR was used to amplify the desired region using a 5' primer that anneals slightly upstream of the SP6 promoter and a 3' primer that anneals at the desired site of truncation. The 3' primer additionally encodes the residues "MLKV" to improve radiolabeling (via the methionine) and stability of the peptidyl-tRNA from hydrolysis during gel electrophoresis (Shao et al., 2013). The PCR products were used in transcription and translation reactions as described above to generate ribosome-nascent chain complexes (RNCs). Following translation, cycloheximide was added to a final concentration of 50µg/ml prior to the addition of membranes. Microsomes were then added as indicated in the figure legends, incubated for 32°C for 15 min, then returned to ice for subsequent protease-protection assays as described above.

Site-specific photo-crosslinking

Incorporation of benzoyl-phenylalanine (Bpa) at specific positions during in vitro translation was accomplished by amber suppression as described (Lin et al., 2019). In short, an amber codon containing template was translated in RRL as above but with 5 µM B. Stearothermophilus $tRNA^{Tyr}$, 0.25 µM Bpa tyrosyl-tRNA synthetase, and 0.1 mM Bpa. UV irradiation was for 12 min with a hand-held UV lamp (ThermoFisher Scientific, cat. #95034) at 254 nm positioned ~2.5 cm above the sample sitting on ice. As described before(Lin et al., 2019), E.coli Bpa tyrosyl-tRNA synthetase was purified via the C-terminal His tag on a 5 mL HiTrap Ni-NTA column (GE), desalted by a gel filtration column on FPLC and concentrated by Amicon Ultra centrifugal filter (Millipore, Z717185-8EA). B. stearothermophilus $tRNA^{Tyr}$, was synthesized by in vitro transcription as before(Lin et al., 2019). The pRSET-based construct encoding the tRNA was digested with BstN1, yielding a DNA fragment

containing the exact $tRNA^{Tyr}$ sequence under a T7 promoter. 5 mL transcription reaction was carried out with 1.2 mg DNA template, 1 mM spermidine, 5 mM DTT, 0.1% Triton, 5 mM NTPs, 25 µM $MgCl^2$, 20 µg/mL E. coli pyrophosphatase, 20 µg/mL T7 polymerase and 125 U Recombinant RNasin (Promega) for 4 hours at 37 °C. The reaction product was digested with Turbo DNase (Ambion) and extracted by acid phenol chloroform extraction to yield purified tRNA.

Cysteine crosslinking

Unless explicitly stated, the crosslinking reactions were performed on isolated RMs. Here, the translation reactions were placed over a 20% sucrose cushion in physiological salt buffer [PSB - 100mM KOAc, 50mM HEPES pH 7.4, 2.5mM $Mg(OAc)_2$], centrifuged in the TLA-55 rotor for 20 min at 4°C, and the pelleted RMs resuspend in one-half the original translation reaction volume of PSB. Bismaleimidohexane (BMH) was added to a final concentration of 250 µM and the reaction was incubated on ice for 1hr to allow the crosslinking reaction to occur. BMH was quenched with 5 mM 2-mercaptoethanol. Aliquots of the reaction were removed at different stage of the process for analysis, as indicated in the individual figure legends. The samples were either used directly for downstream applications (primarily immunoprecipitations as described below) or flash frozen in liquid nitrogen and stored at -80°C for later analysis.

Large scale affinity purification of crosslinked substrate complexes

1 mL translation reactions in the presence of 0.8 mM cold methionine were carried out to produce 1xFLAG tagged or TwinStrep tagged (as a control) Rho TM1+2 146mer RNCs. Membranes were isolated and cysteine-based crosslinking was carried out as described above. Isolated membranes resuspended in PSB were solubilized on ice by addition of an equal volume of 2x solubilization buffer (300 mM KOAc, 2% deoxyBigChap, 20 mM EDTA). RNase A was added to a final concentration of 10 ng/ml and RNA digestion was carried out on ice for 20 min before the solubilised extracts were cleared by centrifugation at 100,000 rpm in the TLA120.2 (Beckman). Cleared extracts were directly immunoprecipitated with FLAG M2 affinity resin (Sigma) in batch at 4°C with end-over-end rotation for 2.5 h. Unbound

fraction was removed, and beads were washed 4 times in 1x NSB [200mM KOAc, 1%DBC, 10mM EDTA, 25mM HEPES, 1mM $Mg(OAc)_2$], then twice in 1x NSB without detergent. Beads were resuspended in 1.5x the bead volume of 200 mM KOAc and 25mM HEPES pH 7.4 for direct trypsinisation and analysis by quantitative mass spectrometry.

Purification of EMC and SRP receptor

SRP receptor (SR) was purified using an affinity resin coupled to anti-SR-alpha as described (Görlich and Rapoport, 1993). The purified SR was then bound to SP-sepharose resin, washed, and detergent-exchanged as described (Görlich and Rapoport, 1993). The final SR preparation is in a buffer containing 50 mM Hepes, pH 7.4, 750 mM KOAc, 5 mM Mg(OAc)2, 0.3% deoxy-BigChap (DBC). EMC was purified as described previously (Guna et al., 2017) and minor contaminants removed by a cation exchange step as follows. Flp-In 293T-Rex cells with stably expressed EMC5-FLAG were induced by the addition of 1µg/mL of doxycycline for 48 hr prior to collection. A ~2.5 g pellet of cells was resuspended in 20 mL of solubilization buffer [50 mM HEPES, 200 mM NaCl, 2 mM $Mg(OAc)_2$, 1% DBC, and EDTA free Protease Inhibitor cocktail (Roche)]. After 30 min on ice, the lysate was cleared by centrifugation at 21,000 x g for 20 min at 4°C in the JA-25.50 rotor (Beckman Coulter). The cleared lysate was then added to 500 µl (packed) of anti-FLAG M2 affinity gel pre-equilibrated in wash buffer 1 [50 mM HEPES, 200 mM NaCl, 2 mM $Mg(OAc)_2$, 0.3% DBC] and incubated at 4°C rotating for 1 hr. The affinity resin was collected by brief centrifugation and washed 5 times in 8 resin volumes of wash buffer 1. EMC was eluted in 1 mL elution buffer [50 mM HEPES, 100 mM NaCl, 2 mM $Mg(OAc)_2$, 0.3% DBC, and 250 µg/mL 3xFLAG peptide] by rotating for 30 min at room temperature. The eluate was then passed through a gravity flow column containing 150 µl (packed) SP-Sepharose Fast-Flow that was pre-equilibrated in wash buffer 2 [50 mM HEPES, 50 mM NaCl 2 mM $Mg(OAc)_2$, 0.3% DBC]. The column was washed 4 times with 10 resin volumes of wash buffer 2, and eluted in 200 µl of 50 mM HEPES, 200 mM NaCl, 2 mM $Mg(OAc)_2$, and 0.25% DBC.

Preparation of Total and Sec Depleted protein extracts.

1 ml of canine rough microsomes (at an absorbance at 280 nm of 50) was diluted in an equal volume of ice-cold 50 mM HEPES, pH 7.4, 250 mM Sucrose, 0.15% DBC. Membranes were collected by centrifugation at 100,000 rpm for 15 min at 4°C in the TL100.3 rotor (Beckman Coulter), resuspended in 1 ml of 400 mM KOAc, 50 mM HEPES, 5 mM $Mg(OAc)_2$, 15% glycerol, and divided in two (samples 1 and 2). Sample 1 was adjusted to 10 mM EDTA 0.8% DBC, while sample 2 was adjusted with 0.8% DBC. After 15min on ice, the samples were centrifuged in the TL120.1 rotor (Beckman Coulter) at 100,000 rpm for 30 min at 4°C to pellet insoluble material and ribosomes/subunits. The supernatant from Sample 1 was saved as the "total ER protein" fraction (550 µl). The supernatant from sample 2, which has now been depleted of ~80% of Sec61 via its ribosome association, was passed sequentially over two gravity flow columns containing 200 µl of protein A resin containing anti-Sec61β antibody pre-equilibrated in extraction buffer. The resulting flow through was collected and saved as the "Sec61-depleted ER protein" fraction (550 µl).

Proteoliposome Reconstitutions

Reconstitutions of proteoliposomes (or matched empty liposomes) were performed with minor modifications of previous methods (Görlich and Rapoport, 1993; Guna et al., 2017) as follows. Purified lipids were obtained from Avanti Polar Lipids and a 20 mg/ml stock solution was prepared as before (Guna et al., 2017) containing Phosphatidyl-choline (PC; from bovine liver), Phosphatidyl-ethanolamine (PE; from bovine liver), and synthetic 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-lissamine rhodamine B (rhPE) in a 8:1.9:0.1 ratio. BioBeads-SM2 (BioRad) were prepared by first wetting them with methanol, then washing extensively with distilled water. After all traces of methanol were removed, the beads were adjusted with water so that the settled beads occupied 50% volume. For use in reconstitutions, the BioBeads were dispensed from this 50% slurry in the desired amount, and the excess liquid was removed by aspiration just before use. The volumes of BioBeads referred to below indicate the packed volume of beads.

For reconstitutions with total and Sec61-depleted ER proteins, the detergent-solubilized preparations from above were supplemented with 850 µg lipids from the prepared 20 mg/ml stock prepared as above. Control liposome reconstitutions contained extraction buffer instead of protein extracts. These mixtures were then added to ~350 µl packed BioBeads (prepared as above) and incubated at 4°C for 18 h with gentle end-over-end mixing. The liquid was separated from the BioBeads, diluted with 4 volumes of ice-cold water, and centrifuged for 45 min at 75,000 rpm in a TL100.3 rotor (Beckman). The pellet was resuspended in 90 µl 100 mM KOAc, 50 mM HEPES pH 7.4, 1 mM $Mg(OAc)_2$, 250 mM sucrose. The rhodamine-labeled PC was used to ensure equal membrane recovery, and protein content was visualized by SDS-PAGE followed by Sypro Ruby staining.

For reconstitutions with purified proteins, purified EMC (or its matched buffer control), purified SR (or its matched buffer control), DBC, and lipids were mixed in a final volume of 90 µl; the final mixture contained 0.52% DBC, 42 mM HEPES, pH 7.4, 333 mM KOAc, 44 mM NaCl, 2.67 mM Mg(OAc)2 and XX pmol EMC and YY pmol SR. This was added to 50 ul of BioBeads (packed volume) and incubated with gentle mixing for 16 h at 4°C. The liquid was separated from the BioBeads, diluted with 10 volumes of ice-cold water, and centrifuged for 45 min at 100,000 rpm in a TL100.3 rotor (Beckman). The pellet was resuspended in 15 µl 100 mM KOAc, 50 mM HEPES pH 7.4, 1 mM $Mg(OAc)_2$, 250 mM sucrose, 1 mM DTT. The rhodamine-labeled PC was used to ensure equal membrane recovery, and protein content was visualized by SDS-PAGE and Sypro Ruby staining. The PLs were used immediately for functional assays without freezing.

Sequence analysis

All GPCRs and tail-anchored membrane proteins were retrieved from the curated and reviewed human Uniprot dataset (UniProt Consortium, 2018). GPCRs containing a signal sequence and tail-anchored proteins destined for mitochondria were manually removed from this set. This left 728 GPCRs and 235 tail-anchored proteins. The TMD regions were taken to be those annotated by Uniprot's automated algorithms. Based on these designations, the length of the translocated domain and the charge within the flanking domains were determined. Relative hydrophobicity

was determined using the transmembrane tendency method (Zhao and London, 2006). The charge difference was calculated using the difference between the C- and N-terminal flanking charges (Hartmann et al., 1989).

Quantitative mass spectrometry using TMT labelling

Proteins samples on beads were reduced with 5 mM DTT at 56°C and alkylated with 10 mM iodoacetamide (IAA) in the dark at 22°C. The alkylation reaction was quenched by the addition of DTT and the samples were digested overnight with trypsin (Promega, 0.1 µg) at 37°C. After digestion, each supernatant was transferred to a fresh Eppendorf tube, the bead samples were extracted once with 50% acetonitrile/ 0.1% TFA and combined with the corresponding supernatant. The peptide mixtures were then partially dried in Speed Vac and desalted using home-made C18 (3M Empore) stage tip filled with 2 µl of poros R3 (Applied Biosystems) resin. Bound peptides were eluted sequentially with 30%, 50% and 80% acetonitrile in 0.1%TFA and lyophilized. Dried peptide mixtures from each condition were re-suspended in 20 ul of 7% MeCN and 1 M triethyl ammonium bicarbonate was added to a final concentration of 200 mM. 0.8 mg of TMT10plex reagents (Thermo Fisher Scientific) was re-constituted in 41 µl anhydrous MeCN. 10 µl of TMT (130C or 131) reagent was added to each peptide mixture and incubated for 1 hr at 20°C. The labelling reactions were terminated by incubation with 2.5 µl of 5% hydroxylamine for 15min. The labelled samples were pooled into one Eppendorf tube and the speed Vac was used to evaporate acetonitrile. Peptides were separated on an Ultimate 3000 RSLC nano System (Thermo Scientific), using an acetonitrile gradient, consisting of buffer A (2% MeCN, 0.1% formic acid) and buffer B (80% MeCN, 0.1% formic acid). Eluted peptides were introduced directly via a nanospray ion source into a Q Exactive Plus hybrid quardrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in standard data dependent mode, performed survey full-scan (MS, m/z = 380-1600) with a resolution of 70000, followed by MS2 acquisitions of the 15 most intense ions with a resolution of 35000 and NCE of 33%. MS target values of 3e6 and MS2 target values of 1e5 were used. The isolation window was set as 0.7 m/z and dynamic exclusion was enabled for 40s. The acquired MS/MS raw files were processed using Proteome Discoverer (version

2.1, Thermo Scientific). MS/MS spectra were searched against a Human Reviewed, UniProt Fasta database (download in 2016), using Mascot (version 2.4, Matrix Science) search engine. Carbamidomethylation of cysteines, TMT6plex (N-term) and TMT6plex (K) were set as fixed modifications, while methionine oxidation and N-terminal acetylation (protein) were selected as variable modifications. For reporter ion quantification, the co-isolation threshold is 30 and average reporter S/N threshold is 10. The output file from Proteome Discoverer, the proteins table was filtered for proteins FDR of 1% and exported as excel files used to produce the plot in Fig. 1c.

Preparation of anti-CCDC47 affinity columns

A 50µl bead volume of protein A agarose was diluted in 750µl PBS, placed in 1 mL Pierce Spin columns, then pre-equilibrated with two 800 µl PBS washes. 100 µg of CCDC47 antibody #1 (A305-100A), CCDC47 antibody #2 (A305-101A), or an anti-HA antibody were diluted in 1 mL of PBS then placed over the prepared protein A columns and allowed to pass through the resin by gravity flow. Flow through was collected and passed over the column a second time. Columns were equilibrated with 800 µl of 0.1 M Na-Borate pH 9.0, then antibodies were conjugated to resin by passing through a 1 mL solution by gravity flow of 0.1 M Na-Borate pH 9.0 containing 5 mg/mL dimethyl pimelimidate (DMP). DMP was quenched with 1 mL of 0.2M ethanolamine pH 8.0 and columns were re-equilibrated in 1x PBS and 0.02% NaN_3 for storage at 4°C until use.

Native affinity purification of CCDC47

The purification was performed at the bench on ice. CCDC47 and HA affinity columns were pre-washed with 300 µl of 0.1 M glycine-HCl pH 2.3, neutralized with 800 µl of 1x PBS, then equilibrated in 800 µl of 1x native solubilization buffer (NSB): 200mM KOAc, 10mM EDTA, 1% DBC, 25mM HEPES pH 7.4. 3 mL of RMs, prepared as described above (at an A_{280} value of 75), were solubilized by addition of an equal volume of 2x NSB. RNase A was added to a final concentration of 10 ng/mL and the samples incubated on ice for 30 min during column preparation. The solubilized extract was cleared by centrifugation at 100,000 rpm in the

TLA100.3 (Beckman) for 1 h at 4°C. Supernatant was removed, divided into 3 equal parts, and passed twice over each column (anti-CCDC47 Ab #1, #2, or an anti-HA control). Each column was washed once with 1 mL of 1x NSB. Recovered proteins were eluted with 300 µl of 0.1 M glycine pH 2.3 and elutions were immediately neutralized with 1 M Tris-Cl pH 8.8. For initial small scale experiments using only CCDC47 antibodies (Fig 2A), the protocol was the exact same except 100µl of starting RMs was used and final elutions were TCA precipitated using standard procedures before loading on a gel.

Identification of Asterix by mass spectrometry

Protein samples eluted from the CCDC47 affinity resin were reduced with DTT and alkylated with iodoacetamide. Because initial efforts to identify the co-purifying 10 kD protein via analysis of tryptic digests failed, we re-did the analysis using other proteases. The samples were digested overnight either with trypsin, Glu-C, chymotrypsin or elastase (Promega). Digest mixtures were acidified with formic acid (FA) and a portion of each of these samples was analyzed by nano-scale capillary LC-MS/MS (Ultimate U3000 HPLC, Thermo Scientific Dionex) at a flow of 300 nL/min. A C18 Acclaim PepMap100 5 µm, 100 µm x 20 mm nanoViper (Thermo Scientific Dionex), trapped the peptides prior to separation on an EASY-Spray columns with an acetonitrile gradient. The Eluted peptides were introduced directly via a EASY-Spray ion source into a Q Exactive mass spectrometer (Thermo Scientific). Data dependent analysis was performed using a resolution of 35,000 for the full MS spectrum, followed by ten MS/MS spectra in the orbitrap. MS spectra were collected over a m/z range of 350-1600 m/z. LC-MS/MS data were searched against the UniProt KB database using Mascot (Matrix Science), with a precursor tolerance of 10 ppm and a fragment ion mass tolerance of 0.1 Da. Two missed enzyme cleavages (or no enzyme for elastase) and variable modifications for oxidized methionine, carbamidomethyl cysteine and pyroglutamic acid, were included. MS/MS data were validated using Scaffold (Proteome Software Inc).

Immunoprecipitations

Unless otherwise indicated, all CCDC47 IPs were performed with CCDC47 antibody #1. For immunoprecipitations under denaturing conditions, samples of interest

were first denatured in 1% SDS and 100 mM Tris pH 8.0 and boiled for 2-5 min. Samples were diluted 10-fold in IP buffer (100 mM NaCl, 50 mM HEPES pH 7.4, 1% TritonX-100) then immunoprecipitated in batch with desired antibodies at 4°C rotating end-over-end for 2-4 h. Unbound supernatant was removed by aspiration and beads were washed 3x in IP buffer before elution in 2x SDS-PAGE sample buffer. Native immunoprecipitation followed similar protocols but were performed in the presence of NSB (200 mM KOAc, 25 mM HEPES pH 7.4, 1% DBC). EDTA and RNase A was added prior to solubilization where RNase treatment is indicated in the figure legends. Pulldowns of His-tagged β_1AR intermediates and full length β_1AR were performed using Ni-NTA affinity resin (Invitrogen) in 1xPBS +250 mM NaCl, 0.5% TritonX-100, 10 mM Imidazole.

Sucrose gradient separation

The products of the quenched crosslinking reaction was solubilized in NSB (200 mM KOAc, 25 mM HEPES pH 7.4, 1% DBC). EDTA and RNase A were added to release the nascent chain from the ribosome and digest the attached tRNA. The sample (typically 10 or 20 µL volume) was loaded onto 200 µL micro sucrose gradient (5-25% sucrose in NSB), centrifuged at 55,000 rpm in the TLS-55 rotor (with suitable adaptors) for 2 h 20 min, and fractionated manually into 11 fractions from the top. The final fraction, which contains aggregates and non-solubilized material, was not analyzed. Fractions 1-10 were analyzed by SDS-PAGE and autoradiography (or Coomassie staining to detect endogenous proteins).

Topology mapping by cysteine accessibility

Asterix-knockout HEK293 T-Rex cells were transfected with constructs encoding human Asterix tagged at the C-terminus with a FLAG tag, including variants lacking all cysteines or containing single cysteines as described in the figure. After 48 h, the cells from each well of a 6-well plate were washed once and collected in ice-cold PBS, pelleted by centrifugation, resuspended on ice in 500 µl permeabilization buffer (100 mM KOAc, 10 mM Tris, pH 7.4, 10 mM $MgCl_2$), and adjusted to 0.01% digitonin. After 5 min on ice, the cells were re-sedimented by centrifugation, washed once in permeabilization buffer lacking digitonin, sedimented again, and

resuspended in 30 µl of PSB on ice. An aliquot was removed for the untreated sample, and the remainder was adjusted to 1 mM final concentration of 5,000 Dalton PEG-maleimide prepared in DMSO. After 1 h on ice, the reaction was quenched with 10 mM 2-mercaptoethanol, the cells recovered by centrifugation, washed once with 600 µl PSB, and finally prepared for analysis by SDS-PAGE and immunoblotting.

SDS-PAGE

All samples were placed in 2.5x SDS-PAGE sample buffer (125mM Tris-HCL, pH6.8, 2.5% SDS, 22.5% glycerol, 0.06% bromophenol blue, 250mM DTT) and boiled at 95°C for 3-5min prior to loading on a gel. Tris-Tricine gels were made up of a 12% acrylamide resolving by adding the following amounts of premade stock solutions: 5.7mls water, 9mls 40% Acrylamide/Bis (Biorad), 15mls 2M Tris -HCl, pH8.45, and 0.3mls 10% SDS. The 4% acrylamide stacking gel was composed of the following amounts of premade stock solutions: : 5.2mls water, 1mls 40% Acrylamide/Bis (Biorad), 3.7mls 2M Tris -HCl, pH8.45, and 0.1mls 10% SDS. Gels were run a 100V for ~90min.

Adelman, M.K., Sabatini, D.D., and Blobel, G. (1973). Ribosome-membrane interaction: Nondestructive disassembly of rat liver rough microsomes into ribosomal and membranous components. J. Cell Biol. *56*, 206–229.

Akopian, D., Shen, K., Zhang, X., and Shan, S. (2013). Signal Recognition Particle: An Essential Protein-Targeting Machine.

Andrews, D.W., Walter, P., and Francisco, S. (1991). A Nascent Membrane Protein Is Located Adjacent to ER Membrane Proteins Throughout Its Integration and Translation. *112*, 809–821.

Argos, P., Rao, J.K.M., and Hargrave, P.A. (1982). Structural Prediction of Membrane-Bound Proteins. Eur. J. Biochem. *128*, 565–575.

Baker, J.A., Wong, W.C., Eisenhaber, B., Warwicker, J., and Eisenhaber, F. (2017). Charged residues next to transmembrane regions revisited: "Positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule." BMC Biol. *15*, 1–29.

Baker, K.P., Schaniel, A., Vestweber, D., and Schatz, G. (1990). A yeast mitochondrial outer membrane protein essential for protein import and cell viability. Nature *348*, 605–609.

Baron, L., Paatero, A.O., Morel, J.-D., Impens, F., Guenin-Macé, L., Saint-Auret, S., Blanchard, N., Dillmann, R., Niang, F., Pellegrini, S., et al. (2016). Mycolactone subverts immunity by selectively blocking the Sec61 translocon. J. Exp. Med. *213*, 2885–2896.

Barz, W.P., and Walter, P. (1999). Two endoplasmic reticulum (ER) membrane proteins that facilitate ER-to-Golgi transport of glycosylphosphatidylinositol-anchored proteins. Mol. Biol. Cell *10*, 1043–1059.

Becker, T., Gebert, M., Pfanner, N., and van der Laan, M. (2009). Biogenesis of mitochondrial membrane proteins. Curr. Opin. Cell Biol. *21*, 484–493.

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. Science *278*, 2123–2126.

Beilharz, T., Egan, B., Silver, P.A., Hofmann, K., and Lithgow, T. (2003). Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in Saccharomyces cerevisiae. J. Biol. Chem. *278*, 8219–8223.

Beltzer, J.P., Fiedler, K., Fuhrer, C., Geffen, I., Handschin, C., Wessels, H.P., and Spiess, M. (1991). Charged residues are major determinants of the transmembrane orientation of a signal-anchor sequence. J. Biol. Chem. *266*, 973–978.

Van den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36–44.

Bircham, P.W., Maass, D.R., Roberts, C. a, Kiew, P.Y., Low, Y.S., Yegambaram, M., Matthews, J., Jack, C. a, and Atkinson, P.H. (2011). Secretory pathway genes assessed by high-throughput microscopy and synthetic genetic array analysis. Mol. Biosyst. *7*, 2589–2598.

Blobel, G. (1980). Intracellular protein topogenesis. Proc. Natl. Acad. Sci. U. S. A. *77*, 1496–1500.

Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes: II. reconstitution of functional rough microsomes from heterologous components. J. Cell Biol. *67*, 852–862.

Blond-Elguindi, S., Cwirla, S.E., Dower, W.J., Lipshutz, R.J., Sprang, S.R., Sambrook, J.F., and Gething, M.J.H. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. Cell *75*, 717–728.

Borgese, N., Mok, W., Kreibich, G., and Sabatini, D.D. (1974). Ribosomal-membrane interaction: In vitro binding of ribosomes to microsomal membranes. J. Mol. Biol. *88*, 559–580.

Bošnjak, I., Bojović, V., Šegvić-Bubić, T.S., and Bielen, A. (2014). Occurrence of protein disulfide bonds in different domains of life: A comparison of proteins from the Protein Data Bank. Protein Eng. Des. Sel. *27*, 65–72.

Brechet, A., Buchert, R., Schwenk, J., Boudkkazi, S., Zolles, G., Siquier-Pernet, K., Schaber, I., Bildl, W., Saadi, A., Bole-Feysot, C., et al. (2017). AMPA-receptor specific biogenesis complexes control synaptic transmission and intellectual ability. Nat. Commun. *8*.

Buchanan, B.B., and Balmer, Y. (2005). REDOX REGULATION: A Broadening Horizon. Annu. Rev. Plant Biol. *56*, 187–220.

Chang, Y.C., and Bowie, J.U. (2014). Measuring membrane protein stability under

native conditions. Proc. Natl. Acad. Sci. U. S. A. *111*, 219–224.

Chartron, J.W., Hunt, K.C.L., and Frydman, J. (2016). Cotranslational signal-independent SRP preloading during membrane targeting. Nature *536*, 224–228.

Chen, G.Q., and Gouaux, E. (1999). Probing the folding and unfolding of wild-type and mutant forms of bacteriorhodopsin in micellar solutions: Evaluation of reversible unfolding conditions. Biochemistry *38*, 15380–15387.

Chi, J.H., Roos, J., and Dean, N. (1996). The OST4 gene of Saccharomyces cerevisiae encodes an unusually small protein required for normal levels of oligosaccharyltransferase activity. J. Biol. Chem. *271*, 3132–3140.

Chitwood, P.J., Juszkiewicz, S., Guna, A., Shao, S., and Hegde, R.S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. Cell *175*, 1507-1519.e16.

Connolly, T., and Gilmore, R. (1989). The signal recognition particle receptor mediates the GTP-dependent displacement of SRP from the signal sequence of the nascent polypeptide. Cell *57*, 599–610.

Conti, B.J., Devaraneni, P.K., Yang, Z., David, L.L., and Skach, W.R. (2015). Cotranslational Stabilization of Sec62/63 within the ER Sec61 Translocon Is Controlled by Distinct Substrate-Driven Translocation Events. Mol. Cell *58*, 269–283.

Costa, E.A., Subramanian, K., Nunnari, J., and Weissman, J.S. (2018). Defining the physiological role of SRP in protein-targeting efficiency and specificity. Science *359*, 689–692.

Crawshaw, S.G., Martoglio, B., Meacock, S.L., and High, S. (2004). A misassembled transmembrane domain of a polytopic protein associates with signal peptide peptidase. Biochem. J. *384*, 9–17.

Crowley, K.S., Reinhart, G.D., and Johnson, A.E. (1993). The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. Cell *73*, 1101–1115.

Crowley, K.S., Liao, S., Worrell, V.E., Reinhart, G.D., and Johnson, A.E. (1994). Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore. Cell *78*, 461–471.

Denzer, a J., Nabholz, C.E., and Spiess, M. (1995). Transmembrane orientation of

signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. EMBO J. *14*, 6311–6317.

Derman, A.I., Puziss, J.W., Bassford, P.J., and Beckwith, J. (1993). A signal sequence is not required for protein export in prlA mutants of Escherichia coli. EMBO J. *12*, 879–888.

Deshaies, R.J., and Schekman, R. (1987). A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum. J. Cell Biol. *105*, 633–645.

Deshaies, R.J., Sanders, S.L., Feldheim, D.A., and Schekman, R. (1991). Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. Nature *349*, 806–808.

Do, H., Falcone, D., Lin, J., Andrews, D.W., and Johnson, A.E. (1996). The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. Cell *85*, 369–378.

Döring, K., Ahmed, N., Riemer, T., Suresh, H.G., Vainshtein, Y., Habich, M., Riemer, J., Mayer, M.P., O'Brien, E.P., Kramer, G., et al. (2017). Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. Cell *170*, 298-311.e20.

Dudek, J., Pfeffer, S., Lee, P.H., Jung, M., Cavalié, A., Helms, V., Förster, F., and Zimmermann, R. (2015). Protein transport into the human endoplasmic reticulum. J. Mol. Biol. *427*, 1159–1175.

Ellis, J. (1988). Proteins as molecular chaperones. Nature *328*, 378–379.

Ellis, R.J., and van der Vies, S.M. (1991). Molecular Chaperones. Annu. Rev. Biochem. *60*, 321–347.

Engelman, D.M., Steitz, T.A., and Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. *15*, 321–353.

Engelman, D.M., Chen, Y., Chin, C.-N.C.-N., Curran, A.R.R., Dixon, A.M., Dupuy, A.D., Lee, A.S., Lehnert, U., Matthews, E.E., Reshetnyak, Y.K., et al. (2003). Membrane protein folding: beyond the two stage model. FEBS Lett. *555*, 122–125.

Fagerberg, L., Jonasson, K., Von Heijne, G., Uhlén, M., and Berglund, L. (2010).

Prediction of the human membrane proteome. Proteomics *10*, 1141–1149.

De Felipe, P., Luke, G.A., Hughes, L.E., Gani, D., Halpin, C., and Ryan, M.D. (2006). E unum pluribus: Multiple proteins from a self-processing polyprotein. Trends Biotechnol. *24*, 68–75.

Feng, Q., and Shao, S. (2018). In vitro reconstitution of translational arrest pathways. Methods *137*, 20–36.

Fons, R.D., Bogert, B.A., and Hegde, R.S. (2003). Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. J. Cell Biol. *160*, 529–539.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Gilmore, R., Blobel, G., and Walter, P. (1982a). Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. J. Cell Biol. *95*, 463–469.

Gilmore, R., Walter, P., and Blobel, G. (1982b). Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor. J. Cell Biol. *95*, 470–477.

Goder, V., Junne, T., and Spiess, M. (2004). Sec61p contributes to signal sequence orientation according to the positive-inside rule. Mol. Biol. Cell *15*, 1470–1478.

Gogala, M., Becker, T., Beatrix, B., Armache, J.-P., Barrio-Garcia, C., Berninghausen, O., and Beckmann, R. (2014). Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. Nature *506*, 107–110.

Görlich, D., and Rapoport, T.A. (1993). Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. Cell *75*, 615–630.

Görlich, D., Prehn, S., Hartmann, E., Kalies, K.U., and Rapoport, T.A. (1992a). A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. Cell *71*, 489–503.

Görlich, D., Hartmann, E., Prehn, S., and Rapoport, T. a (1992b). A protein of the endoplasmic reticulum involved early in polypeptide translocation. Nature *357*,

47–52.

Gu, S., Matta, J.A., Lord, B., Harrington, A.W., Sutton, S.W., Davini, W.B., and Bredt, D.S. (2016). Brain α7 Nicotinic Acetylcholine Receptor Assembly Requires NACHO. Neuron *89*, 948–955.

Guna, A., and Hegde, R.S. (2018). Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. Curr. Biol. *28*, R498–R511.

Guna, A., Volkmar, N., Christianson, J.C., and Hegde, R.S. (2017). The ER membrane protein complex is a transmembrane domain insertase. Science (80-.). *473*, eaao3099.

Haber, E., and Anfinsen, C.B. (1962). Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. J. Biol. Chem. *237*, 1839–1844.

Hainzl, T., Huang, S., Meriläinen, G., Brännström, K., and Sauer-Eriksson, A.E. (2011). Structural basis of signal-sequence recognition by the signal recognition particle. Nat. Struct. Mol. Biol. *18*, 389–391.

Halic, M., Becker, T., Pool, M.R., Spahn, C.M.T., Grassucci, R.A., Frank, J., and Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature *427*, 808–814.

Harrington, S.E., and Ben-Tal, N. (2009). Structural determinants of transmembrane helical proteins. Structure *17*, 1092–1103.

Hartmann, E., Rapoport, T.A., and Lodish, H.F. (1989). Predicting the orientation of eukaryotic membrane-spanning proteins. Proc Natl Acad Sci U S A *86*, 5786–5790.

Hauser, A.S., Chavali, S., Masuho, I., Jahn, L.J., Martemyanov, K.A., Gloriam, D.E., and Babu, M.M. (2017). Pharmacogenomics of GPCR Drug Targets. Cell *172*, 41-43.e19.

Hegde, R.S., and Keenan, R.J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol. *12*, 787–798.

Hegde, R.S., Mastrianni, J.A., Scott, M.R., DeFea, K.A., Tremblay, P., Torchia, M., DeArmond, S.J., Prusiner, S.B., and Lingappa, V.R. (1998). A transmembrane form of the prion protein in neurodegenerative disease. Science *279*, 827–834.

von Heijne, G. (2006). Membrane-protein topology. Nat. Rev. Mol. Cell Biol. *7*, 909–918.

Heinrich, S.U., Mothes, W., Brunner, J., and Rapoport, T.A. (2000). The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. Cell *102*, 233–244.

Helenius, A. (1994). How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. Mol. Biol. Cell *5*, 253–265.

Helenius, A., and Aebi, M. (2004). Roles of N-Linked Glycans in the Endoplasmic Reticulum. Annu. Rev. Biochem. *73*, 1019–1049.

Helenius, A., Trombetta, E.S., Hebert, D.N., and Simons, J.F. (1997). Calnexin, calreticulin and the folding of glycoproteins. Trends Cell Biol. *7*, 193–200.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature *433*, 377–381.

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., and Von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature *450*, 1026–1030.

Hettema, E.H., Erdmann, R., van der Klei, I., and Veenhuis, M. (2014). Evolving models for peroxisome biogenesis. Curr. Opin. Cell Biol. *29*, 25–30.

High, S., Andersen, S.S., Görlich, D., Hartmann, E., Prehn, S., Rapoport, T.A., and Dobberstein, B. (1993a). Sec61p is adjacent to nascent type I and type II signal-anchor proteins during their membrane insertion. J. Cell Biol. *121*, 743–750.

High, S., Martoglio, B., Görlich, D., Andersen, S.S., Ashford, A.J., Giner, A., Hartmann, E., Prehn, S., Rapoport, T.A., and Dobberstein, B. (1993b). Site-specific photocross-linking reveals that Sec61p and TRAM contact different regions of a membrane-inserted signal sequence. J. Biol. Chem. *268*, 26745–26751.

Higy, M., Junne, T., and Spiess, M. (2004). Topogenesis of membrane proteins at the endoplasmic reticulum. Biochemistry *43*, 12716–12722.

Hofmann, N.R., and Theg, S.M. (2005). Chloroplast outer membrane protein targeting and insertion. Trends Plant Sci. *10*, 450–457.

Huang, K.S., Bayley, H., Liao, M.J., London, E., and Khorana, H.G. (1981). Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. J. Biol.

Chem. *256*, 3802–3809.

Illergård, K., Kauko, A., and Elofsson, A. (2011). Why are polar residues within the membrane core evolutionary conserved? Proteins Struct. Funct. Bioinforma. *79*, 79–91.

Imberty, A., and Pérez, S. (1995). Stereochemistry of the N-glycosylation sites in glycoproteins. Protein Eng. Des. Sel. *8*, 699–709.

Imperiali, B., and O'Connor, S.E. (1999). Effect of N-linked glycosylatian on glycopeptide and glycoprotein structure. Curr. Opin. Chem. Biol. *3*, 643–649.

Itakura, E., Zavodszky, E., Shao, S., Wohlever, M.L., Keenan, R.J., and Hegde, R.S. (2016). Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. Mol. Cell *63*, 21–33.

Ito, M., Oiso, Y., Murase, T., Kondo, K., Saito, H., Chinzei, T., Racchi, M., and Lively, M.O. (1993). Possible involvement of inefficient cleavage of preprovasopressin by signal peptidase as a cause for familial central diabetes insipidus. J. Clin. Invest. *91*, 2565–2571.

Itskanov, S., and Park, E. (2019). Structure of the posttranslational Sec protein-translocation channel complex from yeast. Science (80-.). *363*, 84–87.

Itzhak, D.N., Tyanova, S., Cox, J., and Borner, G.H.H. (2016). Global, quantitative and dynamic mapping of protein subcellular localization. Elife *5*, 1–36.

Janda, C.Y., Li, J., Oubridge, C., Hernández, H., Robinson, C. V., and Nagai, K. (2010). Recognition of a signal peptide by the signal recognition particle. Nature *465*, 507–510.

Jaud, S., Fernández-Vidal, M., Nilsson, I.M., Meindl-Beinker, N.M., Hübner, N.C., Tobias, D.J., Von Heijne, G., and White, S.H. (2009). Insertion of short transmembrane helices by the Sec61 translocon. Proc. Natl. Acad. Sci. U. S. A. *106*, 11588–11593.

Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S., et al. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. Science *323*, 1693–1697.

Jungnickel, B., and Rapoport, T.A. (1995). A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. Cell *82*, 261–270.

Junne, T., Schwede, T., Goder, V., and Spiess, M. (2007). Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. J. Biol. Chem. *282*, 33201–33209.

Kalbfleisch, T., Cambon, A., and Wattenberg, B.W. (2007). A bioinformatics approach to identifying tail-anchored proteins in the human genome. Traffic *8*, 1687–1694.

Kalies, K.U., Görlich, D., and Rapoport, T.A. (1994). Binding of ribosomes to the rough endoplasmic reticulum mediated by the Sec61p-complex. J. Cell Biol. *126*, 925–934.

Katz, F.N., Rothman, J.E., Lingappa, V.R., Blobel, G., and Lodish, H.F. (1977). Membrane assembly in vitro: synthesis, glycosylation, and asymmetric insertion of a transmembrane protein. Proc. Natl. Acad. Sci. U. S. A. *74*, 3278–3282.

Keenan, R.J., Freymann, D.M., Walter, P., and Stroud, R.M. (1998). Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell *94*, 181–191.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The Signal Recognition Particle. Annu. Rev. Biochem. *70*, 755–775.

Kida, Y., Morimoto, F., Mihara, K., and Sakaguchi, M. (2006). Function of positive charges following signal-anchor sequences during translocation of the N-terminal domain. J. Biol. Chem. *281*, 1152–1158.

Kim, J., Na, Y.J., Park, S.J., Baek, S.H., and Kim, D.H. (2019). Biogenesis of chloroplast outer envelope membrane proteins. Plant Cell Rep. *38*, 783–792.

Kota, J., and Ljungdahl, P.O. (2005). Specialized membrane-localized chaperones prevent aggregation of polytopic proteins in the ER. J. Cell Biol. *168*, 79–88.

Krishnamani, V., Hegde, B.G., Langen, R., and Lanyi, J.K. (2012). Secondary and tertiary structure of bacteriorhodopsin in the SDS denatured state. Biochemistry *51*, 1051–1060.

Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., et al. (2002). Subcellular localization of the yeast proteome. Genes Dev. *16*, 707–719.

Kundra, R., and Kornfeld, S. (1999). Asparagine-linked oligosaccharides protect Lamp-1 and Lamp-2 from intracellular proteolysis. J. Biol. Chem. *274*, 31039–

31046.

Kutay, U., Hartmann, E., and Rapoport, T.A. (1993). A class of membrane proteins with a C-terminal anchor. Trends Cell Biol. *3*, 72–75.

Lahiri, S., Chao, J.T., Tavassoli, S., Wong, A.K.O., Choudhary, V., Young, B.P., Loewen, C.J.R., and Prinz, W.A. (2014). A Conserved Endoplasmic Reticulum Membrane Protein Complex (EMC) Facilitates Phospholipid Transfer from the ER to Mitochondria. PLoS Biol. *12*.

Lau, F.W., and Bowie, J.U. (1997). A method for assessing the stability of a membrane protein. Biochemistry *36*, 5884–5892.

Lear, J.D., DeGrado, W.F., Choma, C., and Gratkowski, H. (2000). Asparagine-mediated self-association of a model transmembrane helix. Nat. Struct. Biol. *7*, 161–166.

Lee, C.C., and Samuels, E.R. (1964). The kinetics of reaction between L -cysteine hydrochloride and some maleimides. Can. J. Chem. *42*, 168–170.

Li, L., Park, E., Ling, J.J., Ingram, J., Ploegh, H., and Rapoport, T.A. (2016). Crystal structure of a substrate-engaged SecY protein-translocation channel. Nature *531*, 395–399.

Lin, Z., Lin, Z., Gasic, I., Chandrasekaran, V., Peters, N., and Shao, S. (2019). TTC5 mediates autoregulation of tubulin via mRNA degradation. *4352*, 1–9.

Louie, R.J., Guo, J., Rodgers, J.W., White, R., Shah, N., Pagant, S., Kim, P., Livstone, M., Dolinski, K., McKinney, B.A., et al. (2012). A yeast phenomic model for the gene interaction network modulating CFTR-ΔF508 protein biogenesis. Genome Med. *4*, 103.

Lu, P., Min, D., DiMaio, F., Wei, K.Y., Vahey, M.D., Boyken, S.E., Chen, Z., Fallas, J.A., Ueda, G., Sheffler, W., et al. (2018). Accurate computational design of multipass transmembrane proteins. Science (80-.). *359*, 1042–1046.

Luesch, H., and Paavilainen, V.O. (2020). Natural Product Reports Natural products as modulators of eukaryotic protein secretion.

Ma, C., Agrawal, G., and Subramani, S. (2011). Peroxisome assembly: Matrix and membrane protein biogenesis. J. Cell Biol. *193*, 7–16.

MacKinnon, A.L., Paavilainen, V.O., Sharma, A., Hegde, R.S., and Taunton, J. (2014). An allosteric Sec61 inhibitor traps nascent transmembrane helices at the

lateral gate. Elife *2014*, 1–23.

Malkin, L.I., and Rich, A. (1967). Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. J. Mol. Biol. *26*, 329–346.

Martoglio, B., Hofmann, M.W., Brunner, J., and Dobberstein, B. (1995). The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. Cell *81*, 207–214.

Matlack, K.E.S., Misselwitz, B., Plath, K., and Rapoport, T.A. (1999). BIP acts as a molecular ratchet during posttranslational transport of prepro-α factor across the ER membrane [2]. Cell *97*, 553–564.

Mayer, M.P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. Cell. Mol. Life Sci. *62*, 670–684.

McCormick, P.J., Miao, Y., Shao, Y., Lin, J., and Johnson, A.E. (2003). Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. Mol. Cell *12*, 329–341.

McKenna, M., Simmonds, R.E., and High, S. (2017). Mycolactone reveals the substrate-driven complexity of Sec61-dependent transmembrane protein biogenesis. J. Cell Sci. *130*, 1307–1320.

Meacock, S.L., Lecomte, F.J.L., Crawshaw, S.G., and High, S. (2002). Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. Mol. Biol. Cell *13*, 4114–4129.

Meyer, D.I., and Dobberstein, B. (1980). Identification and characterization of a membrane component essential for the translocation of nascent proteins across the membrane of the endoplasmic reticulum. J. Cell Biol. *87*, 503–508.

Milstein, C., Brownlee, G.G., Harrison, T.M., and Mathews, M.B. (1972). A possible precursor of immunoglobulin light chains. Nat. New Biol. *239*, 117–120.

Morimoto, M., Waller-Evans, H., Ammous, Z., Song, X., Strauss, K.A., Pehlivan, D., Gonzaga-Jauregui, C., Puffenberger, E.G., Holst, C.R., Karaca, E., et al. (2018). Bi-allelic CCDC47 Variants Cause a Disorder Characterized by Woolly Hair, Liver Dysfunction, Dysmorphic Features, and Global Developmental Delay. Am. J. Hum. Genet. *103*, 794–807.

Mothes, W., Heinrich, S.U., Graf, R., Nilsson, I., von Heijne, G., Brunner, J., and

Rapoport, T.A. (1997). Molecular mechanism of membrane protein integration into the endoplasmic reticulum. Cell *89*, 523–533.

Mukhopadhyay, R., Ho, Y.S., Swiatek, P.J., Rosen, B.P., and Bhattacharjee, H. (2006). Targeted disruption of the mouse Asna1 gene results in embryonic lethality. FEBS Lett. *580*, 3889–3894.

Müsch, A., Wiedmann, M., and Rapoport, T.A. (1992). Yeast Sec proteins interact with polypeptides traversing the endoplasmic reticulum membrane. Cell *69*, 343–352.

Natarajan, N., Foresti, O., Wendrich, K., Stein, A., and Carvalho, P. (2019). Quality Control of Protein Complex Assembly by a Transmembrane Recognition Factor. Mol. Cell *77*, 108-118.e10.

Navarre, C., Catty, P., Leterme, S., Dietrich, F., and Goffeau, A. (1994). Two distinct genes encode small isoproteolipids affecting plasma membrane H+-ATPase activity of Saccharomyces cerevisiae. J. Biol. Chem. *269*, 21262–21268.

Nilsson, I., and Von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J. Biol. Chem. *268*, 5798–5801.

Noiva, R. (1999). Protein disulfide isomerase: the multifunctional redox chaperone of the endoplasmic reticulum. Semin. Cell Dev. Biol. *10*, 481–493.

Ogg, S.C., and Walter, P. (1995). SRP samples nascent chains for the presenceof signal sequences by interacting with ribosomes at a discrete step during translation elongation. Cell *81*, 1075–1084.

Öjemalm, K., Halling, K.K., Nilsson, I., and Von Heijne, G. (2012). Orientational Preferences of Neighboring Helices Can Drive ER Insertion of a Marginally Hydrophobic Transmembrane Helix. Mol. Cell *45*, 529–540.

Olden, K., Parent, J.B., and White, S.L. (1982). Carbohydrate moieties of glycoproteins a re-evaluation of their function. BBA - Rev. Biomembr. *650*, 209–232.

Oliver, J., Jungnickel, B., Görlich, D., Rapoport, T., and High, S. (1995). The Sec61 complex is essential for the insertion of proteins into the membrane of the endoplasmic reticulum. FEBS Lett. *362*, 126–130.

Osborne, R.S., and Silhavy, T.J. (1993). PrlA suppressor mutations cluster in

regions corresponding to three distinct topological domains. EMBO J. *12*, 3391–3398.

Ott, C.M., and Lingappa, V.R. (2002). Integral membrane protein biosynthesis: why topology is hard to predict. J. Cell Sci. *115*, 2003–2009.

Paetzel, M., Dalbey, R.E., and Strynadka, N.C.J. (1998). Crystal structure of a bacterial signal peptidase in complex with a β-lactam inhibitor. Nature *396*, 186–190.

Panzner, S., Dreier, L., Hartmann, E., Kostka, S., and Rapoport, T.A. (1995). Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. Cell *81*, 561–570.

Pelham, H.R.B. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell *46*, 959–961.

Perlman, J.H., Wang, W., Nussenzveig, D.R., and Gershengorn, M.C. (1995). A disulfide bond between conserved extracellular cysteines in the thyrotropin-releasing hormone receptor is critical for binding. J. Biol. Chem. *270*, 24682–24685.

Petrescu, A.J., Milac, A.L., Petrescu, S.M., Dwek, R.A., and Wormald, M.R. (2004). Statistical analysis of the protein environment of N-glycosylation sites: Implications for occupancy, structure, and folding. Glycobiology *14*, 103–114.

Plath, K., and Rapoport, T.A. (2000). Spontaneous release of cytosolic proteins from posttranslational substrates before their transport into the endoplasmic reticulum. J. Cell Biol. *151*, 167–178.

Plath, K., Mothes, W., Wilkinson, B.M., Stirling, C.J., and Rapoport, T.A. (1998). Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. Cell *94*, 795–807.

Popot, J.L., and Engelman, D.M. (1990). Membrane Protein Folding and Oligomerization: The Two-Stage Model. Biochemistry *29*, 4031–4037.

Popot, J.L., and Engelman, D.M. (2000). Helical membrane protein folding, stability, and evolution. Annu. Rev. Biochem. *69*, 881–922.

Powers, T., and Walter, P. (1995). Reciprocal stimulation of GTP hydrolysis by two directly interacting GTpases. Science (80-.). *269*, 1422–1424.

Racchi, M., Watzke, H.H., High, K.A., and Lively, M.O. (1993). Human

coagulation factor X deficiency caused by a mutant signal peptide that blocks cleavage by signal peptidase but not targeting and translocation to the endoplasmic reticulum. J. Biol. Chem. *268*, 5735–5740.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663–669.

Reddy, P.S., and Corley, R.B. (1998). Assembly, sorting, and exit of oligomeric proteins from the endoplasmic reticulum. BioEssays *20*, 546–554.

Redman, C.M., and Sabatini, D.D. (1966). Vectorial discharge of peptides released by puromycin from attached ribosomes. Proc. Natl. Acad. Sci. U. S. A. *56*, 608–615.

Reinders, J., Zahedi, R.P., Pfanner, N., Meisinger, C., and Sickmann, A. (2006). Toward the complete yeast mitochondrial proteome: Multidimensional separation techniques for mitochondrial proteomics. J. Proteome Res. *5*, 1543–1554.

Richard, M., Boulin, T., Robert, V.J.P., Richmond, J.E., and Bessereau, J.-L. (2013). Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. Proc. Natl. Acad. Sci. U. S. A. *110*, E1055-63.

Rosenbaum, D.M., Rasmussen, S.G.F., and Kobilka, B.K. (2009). The structure and function of G-protein-coupled receptors. Nature *459*, 356–363.

Rothblatt, J.A., Deshaies, R.J., Sanders, S.L., Daum, G., and Schekman, R. (1989). Multiple genes are required for proper insertion of secretory proteins into the endoplasmic reticulum in yeast. J. Cell Biol. *109*, 2641–2652.

Rothe, C., and Lehle, L. (1998). Sorting of invertase signal peptide mutants in yeast dependent and independent on the signal-recognition particle. Eur. J. Biochem. *252*, 16–24.

Sabatini, D.D., and Blobel, G. (1970). Controlled proteolysis of nascent polypeptides in rat liver cell fractions: II. Location of the polypeptides in rough microsomes. J. Cell Biol. *45*, 146–157.

Sadlish, H., Pitonzo, D., Johnson, A.E., and Skach, W.R. (2005). Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. Nat. Struct. Mol. Biol. *12*, 870–878.

Sakaguchi, M., Mihara, K., and Sato, R. (1987). A short amino-terminal segment

of microsomal cytochrome P-450 functions both as an insertion signal and as a stop-transfer sequence. EMBO J. *6*, 2425–2431.

Saksena, S., Shao, Y., Braunagel, S.C., Summers, M.D., and Johnson, A.E. (2004). Cotranslational integration and initial sorting at the endoplasmic reticulum translocon of proteins destined for the inner nuclear membrane. Proc. Natl. Acad. Sci. U. S. A. *101*, 12537–12542.

Sanders, S.L., Whitfield, K.M., Vogel, J.P., Rose, M.D., and Schekman, R.W. (1992). Sec61p and BiP directly facilitate polypeptide translocation into the ER. Cell *69*, 353–365.

Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., et al. (2017). A comprehensive map of molecular drug targets. Nat. Rev. Drug Discov. *16*, 19–34.

Sato, B.K., Schulz, D., Do, P.H., and Hampton, R.Y. (2009). Misfolded Membrane Proteins Are Specifically Recognized by the Transmembrane Domain of the Hrd1p Ubiquitin Ligase. Mol. Cell *34*, 212–222.

Sato, K., Sato, M., and Nakano, A. (2003). Rer1p, a Retrieval Receptor for ER Membrane Proteins, Recognizes Transmembrane Domains in Multiple Modes. Mol. Biol. Cell *14*, 3605–3616.

Satoh, T., Ohba, A., Liu, Z., Inagaki, T., and Satoh, A.K. (2015). dPob/EMC is essential for biosynthesis of rhodopsin and other multi-pass membrane proteins in Drosophila photoreceptors. Elife *87*, 781–784.

Savidis, G., Mcdougall, W.M., Meraner, P., Green, S., Kowalik, T.F., Brass Correspondence, A.L., Perreira, J.M., Portmann, J.M., Trincucci, G., John, S.P., et al. (2016). Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics. Cell Rep. *16*, 232–246.

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P., Koerten, H.K., Koning, R.I., and Ban, N. (2006). Structure of the E. coli signal recognition particle bound to a translating ribosome. Nature *444*, 503–506.

Schibich, D., Gloge, F., Pöhner, I., Björkholm, P., Wade, R.C., Von Heijne, G., Bukau, B., and Kramer, G. (2016). Global profiling of SRP interaction with nascent polypeptides. Nature *536*, 219–223.

Schlüter, A., Fourcade, S., Domènech-Estévez, E., Gabaldón, T., Huerta-Cepas, J.,

Berthommier, G., Ripp, R., Wanders, R.J.A., Poch, O., and Pujol, A. (2007). PeroxisomeDB: A database for the peroxisomal proteome, functional genomics and disease. Nucleic Acids Res. *35*, 815–822.

Schrul, B., Kapp, K., Sinning, I., and Dobberstein, B. (2010). Signal peptide peptidase (SPP) assembles with substrates and misfolded membrane proteins into distinct oligomeric complexes. Biochem. J. *427*, 523–534.

Sevier, C.S., and Kaiser, C.A. (2002). Formation and transfer of disulphide bonds in living cells. Nat. Rev. Mol. Cell Biol. *3*, 836–847.

Shao, S., and Hegde, R.S. (2011). Membrane protein insertion at the endoplasmic reticulum. Annu. Rev. Cell Dev. Biol. *27*, 25–56.

Shao, S., Von der Malsburg, K., and Hegde, R.S. (2013). Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. Mol. Cell *50*, 637–648.

Sharma, A., Mariappan, M., Appathurai, S., and Hegde, R.S. (2010). In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. Methods Mol. Biol. *619*, 339–363.

Sharpe, H.J., Stevens, T.J., and Munro, S. (2010). A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties. Cell *142*, 158–169.

Shurtleff, M.J., Itzhak, D.N., Hussmann, J.A., Schirle Oakdale, N.T., Costa, E.A., Jonikas, M., Weibezahn, J., Popova, K.D., Jan, C.H., Sinitcyn, P., et al. (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. Elife *7*, e37018.

Simon, S.M., and Blobel, G. (1991). A protein-conducting channel in the endoplasmic reticulum. Cell *65*, 371–380.

Smith, M.A., Clemons, W.M., DeMars, C.J., and Flower, A.M. (2005). Modeling the effects of prl mutations on the Escherichia coli SecY complex. J. Bacteriol. *187*, 6454–6465.

Sonnabend, A., Spahn, V., Stech, M., Zemella, A., Stein, C., and Kubick, S. (2017). Production of G protein-coupled receptors in an insect-based cell-free system. Biotechnol. Bioeng. *114*, 2328–2338.

Spiess, M., and Lodish, H.F. (1986). An internal signal sequence: The

asialoglycoprotein receptor membrane anchor. Cell *44*, 177–185.

Stein, K.C., Kriel, A., and Frydman, J. (2019). Nascent Polypeptide Domain Topology and Elongation Rate Direct the Cotranslational Hierarchy of Hsp70 and TRiC/CCT. Mol. Cell *75*, 1117-1130.e5.

Stirling, C.J., Rothblatt, J., Hosobuchi, M., Deshaies, R., and Schekman, R. (1992). Protein translocation mutants defective in the insertion of integral membrane proteins into the endoplasmic reticulum. Mol. Biol. Cell *3*, 129–142.

Talbot, B.E., Vandorpe, D.H., Stotter, B.R., Alper, S.L., and Schlondorff, J.S. (2019). Transmembrane insertases and N -glycosylation critically determine synthesis, trafficking, and activity of the nonselective cation channel TRPC6 . J. Biol. Chem. *294*, 12655–12669.

Trombetta, E.S. (2003). The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. Glycobiology *13*, 77–91.

Trueman, S.F., Mandon, E.C., and Gilmore, R. (2012). A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. J. Cell Biol. *199*, 907–918.

Tsai, B., Rodighiero, C., Lencer, W.I., and Rapoport, T.A. (2001). Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin. Cell *104*, 937–948.

Ulbrandt, N.D., Newitt, J.A., and Bernstein, H.D. (1997). The E. coli signal recognition particle is required for the insertion of a subset of inner membrane proteins. Cell *88*, 187–196.

Ulmschneider, M.B., Sansom, M.S.P., and Di Nola, A. (2005). Properties of integral membrane protein structures: Derivation of an implicit membrane potential. Proteins Struct. Funct. Genet. *59*, 252–265.

UniProt Consortium, T. (2018). UniProt: the universal protein knowledgebase. Nucleic Acids Res. *46*, 2699–2699.

UniProt Consortium, T. (2018). Erratum: UniProt: the universal protein knowledgebase (Nucleic acids research (2017) 45 D1 (D158-D169)). Nucleic Acids Res. *46*, 2699.

Valent, Q.A., Kendall, D.A., High, S., Kusters, R., Oudega, B., and Luirink, J. (1995). Early events in preprotein recognition in E. coli: interaction of SRP and

trigger factor with nascent polypeptides. EMBO J. *14*, 5494–5505.

Venkatakrishnan, A.J., Deupi, X., Lebon, G., Tate, C.G., Schertler, G.F., and Madan Babu, M. (2013). Molecular signatures of G-protein-coupled receptors. Nature *494*, 185–194.

Vinothkumar, K.R., and Henderson, R. (2010). Structures of membrane proteins. Q. Rev. Biophys. *43*, 65–158.

Voorhees, R.M., and Hegde, R.S. (2015). Structures of the scanning and engaged states of the mammalian srp-ribosome complex. Elife *4*, 1–21.

Voorhees, R.M., and Hegde, R.S. (2016). Structure of the Sec61 channel opened by a signal sequence. Science *351*, 88–91.

Voorhees, R.M., Fernández, I.S., Scheres, S.H.W., and Hegde, R.S. (2014). Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. Cell *157*, 1632–1643.

Wahlberg, J.M., and Spiess, M. (1997). Multiple determinants direct the orientation of signal-anchor proteins: The topogenic role of the hydrophobic signal domain. J. Cell Biol. *137*, 555–562.

Walter, P., and Blobel, G. (1980). Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. Proc. Natl. Acad. Sci. U. S. A. *77*, 7112–7116.

Walter, P., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. J. Cell Biol. *91*, 557–561.

Walter, P., and Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature *299*, 691–698.

Walter, P., and Blobel, G. (1983). Preparation of microsomal membranes for cotranslational protein translocation. Methods Enzymol. *96*, 84–93.

Walter, P., Jackson, R.C., Marcus, M.M., Lingappa, V.R., and Blobel, G. (1979). Tryptic dissection and reconstitution of translocation activity for nascent presecretory proteins across microsomal membranes. Proc. Natl. Acad. Sci. U. S. A. *76*, 1795–1799.

Wang, Q., and Chang, A. (2003). Substrate recognition in ER-associated degradation mediated by Eps1, a member of the protein disulfide isomerase family. EMBO J. *22*, 3792–3802.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science (80-.). *350*, 1096–1101.

Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G.W., Tate, C.G., and Schertler, G.F.X. (2008). Structure of a β1-adrenergic G-protein-coupled receptor. Nature *454*, 486–491.

Warne, T., Serrano-Vega, M.J., Tate, C.G., and Schertler, G.F.X. (2009). Development and crystallization of a minimal thermostabilised G protein-coupled receptor. Protein Expr. Purif. *65*, 204–213.

Wawrzynow, A., Theibert, J.L., Murphy, C., Jona, I., Martonosi, A., and Collins, J.H. (1992). Sarcolipin, the "proteolipid" of skeletal muscle sarcoplasmic reticulum, is a unique, amphipathic, 31-residue peptide. Arch. Biochem. Biophys. *298*, 620–623.

Wessels, H.P., and Spiess, M. (1988). Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence. Cell *55*, 61–70.

Wiedmann, M., Kurzchalia, T. V., Hartmann, E., and Rapoport, T.A. (1988). A signal sequence receptor in the endoplasmic reticulum membrane. Nature *328*, 830–833.

Wormald, M.R., and Dwek, R.A. (1999). Glycoproteins: Glycan presentation and protein-fold stability. Structure *7*, 155–160.

Wormald, M.R., Petrescu, A.J., Pao, Y.L., Glithero, A., Elliott, T., and Dwek, R.A. (2002). Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling. Chem. Rev. *102*, 371–386.

Wu, X., Cabanos, C., and Rapoport, T.A. (2019). Structure of the post-translational protein translocation machinery of the ER membrane. Nature *566*, 136–139.

Yamamoto, S., Yamazaki, T., Komazaki, S., Yamashita, T., Osaki, M., Matsubayashi, M., Kidoya, H., Takakura, N., Yamazaki, D., and Kakizawa, S. (2014). Contribution of calumin to embryogenesis through participation in the

endoplasmic reticulum-associated degradation activity. Dev. Biol. *393*, 33–43.

Yang, Y., Guo, R., Gaffney, K., Kim, M., Muhammednazaar, S., Tian, W., Wang, B., Liang, J., and Hong, H. (2018). Folding-Degradation Relationship of a Membrane Protein Mediated by the Universally Conserved ATP-Dependent Protease FtsH. J. Am. Chem. Soc. *140*, 4656–4665.

Yernool, D., Boudker, O., Jin, Y., and Gouaux, E. (2004). Structure of a glutamate transporter homologue from Pyrococcus horikoshii. Nature *431*, 811–818.

Zhang, Z.R., Bonifacino, J.S., and Hegde, R.S. (2013). Deubiquitinases sharpen substrate discrimination during membrane protein degradation from the ER. Cell *154*, 609–622.

Zhao, G., and London, E. (2006). An amino acid "transmembrane tendency" scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. Protein Sci. *15*, 1987–2001.

Zhou, F.X., Cocco, M.J., Russ, W.P., Brunger, A.T., and Engelman, D.M. (2000). Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nat. Struct. Biol. *7*, 154–160.

Zopf, D., Bernstein, H.D., Johnson, A.E., and Walter, P. (1990). The methionine-rich domain of the 54 kd protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. EMBO J. *9*, 4511–4517.